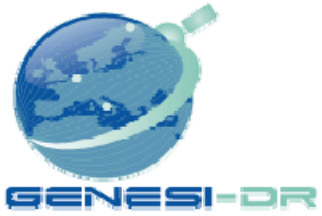


# DNA4.2

## Description of European Archiving Infrastructure

| | | | |
|---|---|---|---|
| **Activity:** | NA4 | Data and Resources Access Policies | |
| **Task:** | TNA4.3 | Development of a sustainable European Archiving Infrastructure concept | |
| **Author(s):** | Cornelia Boer, Frederic Berriri Francesco Coccoluto, | | CS SI |
| | Andrea Kaiser-Weiss, Adit Santokhee | | RUR |
| | Klaus-Dieter Missling | | DLR |
| | Giuseppe Manzella | | ENEA |
| | Thomas Lankester | | INFOTERRA |
| | Jean-Marie Wallut | | CNES |
| | Simone Gentilini | | JRC |
| | Marco Serra, Luigi Dini | | ASI |
| | Roberto Cossu, Eliana Li Santi | | ESA |
| **Authorized by** | Klaus-Dieter Missling | | DLR |
| **Doc Id:** | GENESI-DR-NA4-DEL-DNA4.2 | | |
| **Reviewer** | Fabrizio Pacini | | ED |
| **Reviewer** | Joost van Bemmelen | | ESA |
| **Dissemination Level** | Restricted to other programme participants | | |

**Abstract:**
This document describes the existing and planned European infrastructures on data archive and dissemination.

# Document Log

| Date | Author | Changes | Version | Status |
|---|---|---|---|---|
| 11/08/2008 | F.Coccoluto | Document structure | 0.5 | Draft |
| 25/09/2008 | F.Coccoluto<br>A.Kaiser-Weiss | Updates | 0.6 | QA Draft |
| 21/10/2009 | C. Boer | Finalization of the document, taking into account the internal reviewers' comments. | 0.7 | Delivery Candidate |
| 03/11/2008 | C. Boer | Approved by GPEB | 1.0 | Delivered |
| 29/05/2009 | C. Boer | Initialisation of final document | 1.1 | Draft |
| 12/06/2009 | A.Kaiser-Weiss | EUMETSAT, EUSC, ICCM, INFOTERRA, RISK-EOS chapters | 1.2 | Draft |
| 15/06/2009 | K.-D. Missling | BSH, DLR, DLS, ETH, EURAC, MRI, NERSC, NILU, NINA, WDC-Climate, ZALF, Planetek Data Processing Centre, PREVIEW Data Repositories, Starlab Data Repository chapters | 1.3 | Draft |
| 17/06/2009 | G.M.R. Manzella | ENEA chapter | 1.4 | Draft |
| 17/06/2009 | Simone Gentilini | JRC chapter | 1.5 | Draft |
| 03/07/2009 | Adit Santokhee | SWISSTOPO, TNO and University of Helsinki chapters | 1.6 | Draft |
| 24/07/2009 | J.M. Wallut | CNES chapter | 1.7 | Draft |
| 24/07/2009 | M. Serra,<br>L. Dini | ASI chapter | 1.8 | Draft |
| 24/07/2009 | F. Berriri | Section 1, CLS, CNIG, INGV, KSAT, NLS, NOA, OGS chapters. | 1.9 | Draft |
| 29/07/2009 | R. Cossu | ESA chapter | 1.10 | Draft |
| 30/07/2009 | C. Boer | Finalization | 1.11 | QA Draft |
| 14/08/2009 | C. Boer | Taking into account internal reviewers' comments; restructuring. | 1.12 | QA Draft |
| 18/02/2010 | E. Li Santi | 3.2 Interoperability and standardisation efforts in Earth Sciences | 1.13 | Delivery Candidate |
| 26/03/2010 | C. Boer<br>K-D. Missling | Approved by GPEB | 2.0 | Delivered |

# Executive Summary

This document provides a map of the existing and planned European infrastructure on data archive and dissemination, based on the experiences of various technical operators at both national and European level.

These experiences have been collected over the period April 2008 to April 2009 via a web-based survey with call for participation sent via e-mail to several partners of the previous and current GMES and INSPIRE projects. The collected feedback has been integrated with the provided reference documentation and the information available in the relevant web sites.

ES needs in terms of infrastructure have been analyzed and infrastructures and infrastructure projects have been identified so as to derive a description of the characteristics of the infrastructures of the coming years. This has enabled to assess GENESI-DR versus the current axes of development in infrastructures.

The document also describes European and national digital repositories (archives) which have answered to the questionnaire in terms of:

- Archive systems: archive backup and recovery systems; archive media in use (transcription systems and policy);
- Access to archive via gateways: connection to high speed networks;
- Contents of the archives (EO missions, duration, coverage and volume), data curation aspects and preparation to long term preservation,

So as to derive the main characteristics of the current digital repositories.

This document is an output of the work package TNA4.3. A first version was provided at milestone M10. It is now enriched by further collection of information, in particular about the DR integrated in GENESI-DR, and by general improvements to make clearer the approach of this document and the usefulness of the derived information.

# TABLE OF CONTENTS

# 1. Introduction

## 1.1 Purpose and Scope

The aim of this document is: to make a survey of the existing and planned infrastructure on data archive and dissemination of Europe.

International organization (such as UNESCO) are encouraging institutions of the Member States to provide timely, free and unrestricted international exchange of data for a wide variety of purposes including the prediction of weather and climate, the preservation of life, the mitigation of human-induced changes in the environment, as well as for the advancement of scientific understanding that makes this possible.

Furthermore, these international organizations are through technical bodies also co-operating with data contributors to ensure that data can be accepted into the appropriate systems and can meet quality requirements. The requirements that can be extracted from these strong demands of the international organizations are:

- Data collection, preprocessing, processing to assure high quality,
- Timely access to data,
- Development, implementation, maintenance of archives of data and related information,
- Access to archives.

These requirements are defining the logical links between Digital Repositories and Infrastructures, data access and curation.

With respect to access and sharing of data, the current European political context is largely determined by:

- The Aarhus Convention [15], which requires public bodies in countries which are parties to share environmental information if asked to do so and to actively disseminate environmental information in their possession.
- The INSPIRE directive [16], which entered into force on 15 May 2007, and which is to pave the way for more efficient geo-spatial data access and sharing between public authorities by gradually removing technical obstacles related to the interoperability of data and through the implementation of data services allowing to find, view and share data from various sources. In addition, Member States will implement data and charging policies so that spatial information needed for good governance is plentiful and seamless available under terms that facilitate its extensive use. At present, the Commission and EEA, in continuous interchange with the relevant stakeholder groups, are proceeding to develop the implementation rules which will aim to make existing spatial data and services interoperable.

GMES will help to identify and consolidate relevant in situ infrastructure and elaborated products in the Member States as a specific function of the data requirements of GMES Core Services. Hence, this presents an important opportunity to improve the synergy and efficiency of European data collection and information capacity in line with the political and legislative framework, clarifying where further efforts by Member States might be needed, and where existing monitoring may be less effective and could therefore be reduced or streamlined.

The provision of in situ data and products to the GMES services should be based on the principle of subsidiarity and make best use, to the maximum extent possible, of existing capacities in Europe, especially those of existing European Agencies. Regarding the

**observation infrastructure**, this provides **opportunities to streamline and focus the resources already deployed in Member States**, based on the existing relevant in situ networks established in the context of obligations from existing legislation (regional, national, European), international commitments and treaties as well as national research efforts. Hence, a part of the contribution of Member States to the provision of GMES services will be to assure the required **continuity of observation**. Another important issue lies with standardisation at European level of the measurement systems and survey characteristics and performances: this requires an **improved coordination of observation systems at EU level**, built on already existing European and international coordination mechanisms.

GENESI-DR (Ground European Network for Earth Science Interoperations – Digital Repositories) has the challenge of establishing open Earth Science Digital Repository access for European and world-wide science users to seamlessly access and share all data, information, products and knowledge originating from space, airborne and in-situ sensors.

GENESI-DR builds upon the existing, operational and focused Earth Observation (EO) European infrastructure and involves key Earth Science centres responsible for operational data acquisition, processing, archiving and distribution. Network activity 4 (NA4) deals with Data and Resources Access Policies to get the project objectives. Therefore on development level it is important to have a list of assessed entities (of European EO infrastructure) which are candidates for a “GENESI-fication”. On the operational level it is important to establish working relations with relevant bodies for coordination and harmonization of data policy issues. A map of the existing and planned European infrastructure on data archive and dissemination was identified as a useful tool for current and future integration of digital repositories. This map should be as complete as possible and contain an assessment of the suitability for GENESI-DR integration.

For that reason the purpose of this document is to provide a map of the existing and planned European infrastructure on data archive and dissemination, based on the experiences of various technical operators at national and European level, to maintain a coherent approach in generalizing the archive base infrastructure for the Earth Science user community.

Therefore the scope of the mapping is European and national Earth Science data infrastructures and digital repositories as pre-condition for an efficient integration into the GENESI-DR (or comparable) infrastructure.

## 1.2 Document Organisation

This document is structured in the following sections:

Section 1 describes the purpose and the scope of the document and provides an overview for the rest of the document.

Section 2 presents the followed approach for collecting the information necessary for providing a map of the European archiving infrastructure.

Section 3 presents existing and planned (i.e. still a project) archive infrastructures at European level and national level.

Section 4 presents the data repositories of the contacted persons who have answered the questionnaire as well as the data repositories of the partners partaking to GENESI-DR Infrastructure (ESA, CNES, DLR, NILU, KSAT, JRC, ENEA-CNR, ASI and INFOTERRA).

Section 5 outlines the main conclusions about the status of European infrastructures and data repositories and the advantages of infrastructures such as GENESI-DR.

Annex C supports the reader by means of Definitions of Terms, Acronyms and Abbreviations.

Annex D contains both the submitted questionnaire and the list of the contacted organisations.

## 1.3 Changes to Previous Version

The first version of DNA4.2 was delivered in M10 and was considered as a draft document, taking into account the answers provided from April to July 2008.

This second version corresponds to the final version of the document. The main changes are:

- The inclusion of a critical review of infrastructure projects;
- The collecting of the second wave of answers from January to April 2009;
- The description of GENESI-DR partners' repositories and infrastructures;
- A re-organization of the contents of the document, which is now divided into two parts: one dedicated to the infrastructures and the second one dedicated to the data repositories.

# 2. Approach

## 2.1 Introduction

The objective of this deliverable is to provide a map of the European Earth Science infrastructures and digital repositories. In order to achieve this objective, there have been two approaches:

- Collection of the information via web-browsing (see Section 2.2);
- Collection of the information via a dedicated questionnaire (see Section 2.3).

This deliverable was used as a means for work in network activity 4 (NA4) especially sub-task *TNA4.3 – Development of a sustainable European Archiving Infrastructure concept*. The overall context is shown in Figure 1. DNA4.2 [12] served as an input for DNA4.4 [13]. It provided entities with noteworthy concepts of a sustainable archive. In the second step DNA4.4 [13] and DSA2.4 [14] provides qualitative assessment criteria used in version 2.

This document serves as a GENESI-DR-internal support survey because a more broader approach requires active and direct cooperation with other initiatives which has to result into agreed questionnaires and assessment issues. The conclusion of this procedure was not realistic for a 2-years project. But for a continued GENESI-DR an assessed map for next DR- inclusion is needed.

**In clear words – in the first place customer of this document is GENESI-DR.** The assessment of European Archiving Infrastructure elements was done vs. GENESI-DR because:

- GENESI-DR has no permission/order from other initiatives,
- GENESI-DR has no special inquiries from such initiatives,
- For a real assessment for one of the initiatives a working contact to this initiative has to be established, a questionnaire with agreed questions has to be used -> this was not realizable in the planned time frame.
- But a continued GENESI-DR will need a assessed map for next DR- inclusion.

Therefore the analysis of European infrastructure (elements of the map) was done in particular versus:

1. ESDR-Minimum-Requirements-and-GENESI-fication-Guide.doc [14]
2. GENESI-DR: concepts of sustainability (DNA4.4), Section 4 European Archive Sustainability Concept [13]

   A sustainable archive is an archive which can be maintained over the long term, i.e. it implies that: there are: Sustainable data, Sustainable data formats, Sustainable metadata formats, Sustainable knowledge base, Sustainable added-value services and Sustainable architecture. Preservation/curation and sustainability aspects are taken in consideration in the questionnaire (D.1, questions 43-59), and standards for metadata, services, delivery … (D.1, questions 36-42).

An additional but only qualitative assessment was done for:

3. INSPIRE (services and metadata)

because:

- INSPIRE is in sense of GENESI-DR- it is considered in the GENESI- design,
- It is an European initiative of high importance and broad application.

Qualitative assessment means - one can see whether the DR are “close” to INSPIRE, i.e. use of ISO19115/ISO19139/19119. This document does not certify any compliance because there was neither a profound assessment of metadata sets nor of implementation of services based on WCS/WFS.

This map can be considered as complementary to the map of European Spatial Data Infrastructures performed in the framework of INSPIRE (see State of play 2007 describing National SDIs in 32 countries in Europe (see reference [2])), as it focuses on Earth Science infrastructures, but there is no link with the approach followed by INSPIRE teams (document organization, kind of information collected).

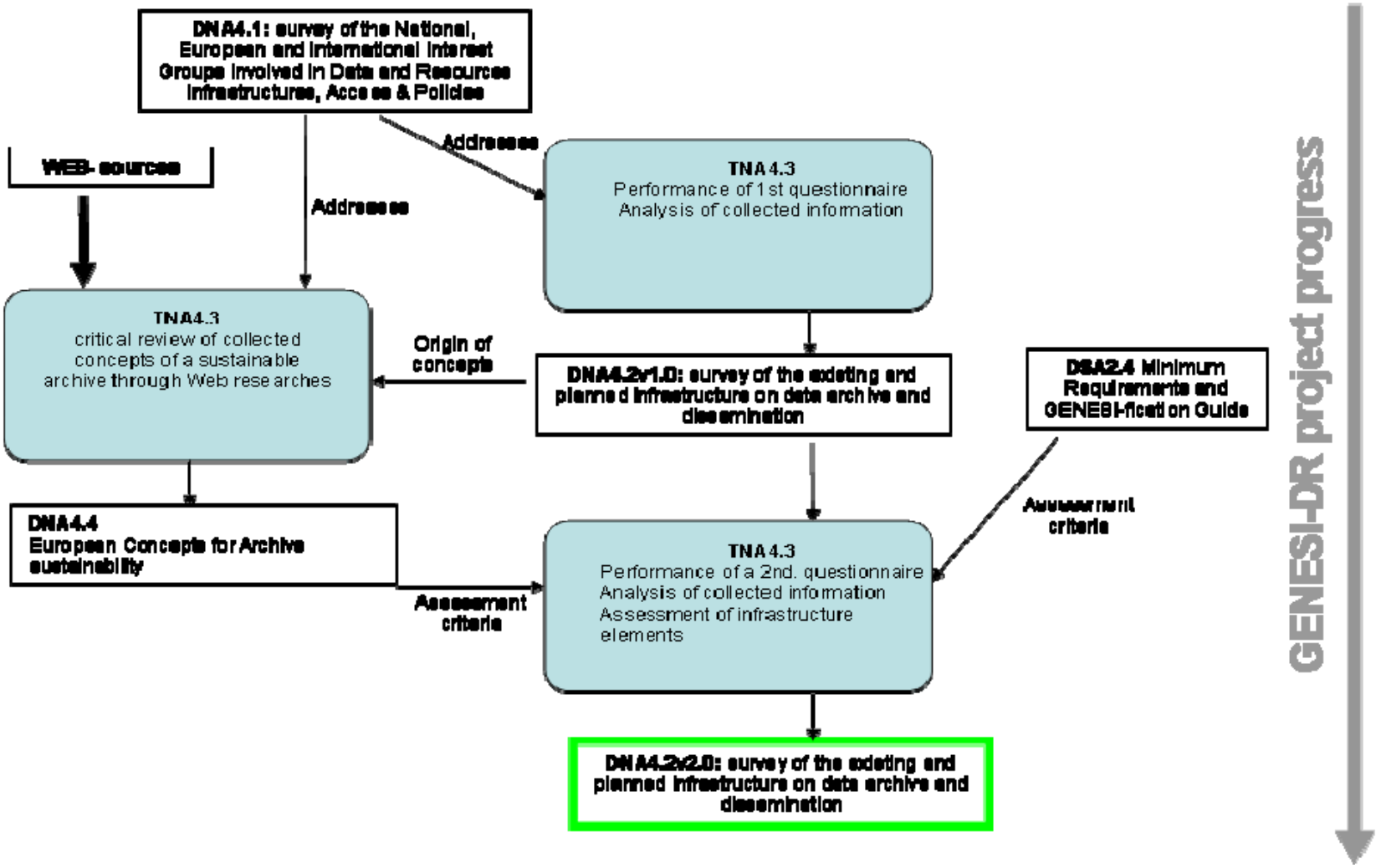


**Figure 1 DNA4.2 - overall context**

## 2.2 Web browsing

For the description of European or national infrastructures (see Section 3), the information provided on the associated Websites has been summarized in the dedicated section.

Please refer to section Annex C for the definition of terms like “infrastructure” which has been considered in this document.

## 2.3 Web-base survey

For the description of existing European digital repositories (see Section 4), the information provided comes from a web-based survey.

Please refer to section Annex C for the definition of terms like “digital/data repository” which has been considered in this document.

GENESI-DR Project organised a web-based survey over the two following periods: from April to July 2008 and from January to April 2009.

The survey was launched via an e-mail invitation addressed to a list of organisations (see D.2) retrieved from previous and current GMES and INSPIRE projects partnerships. An intermediate revision of the contact list was performed after the first month based on the feedback of the contacted people.

The submitted questionnaire (see D.1) focused on the technical infrastructure such as:

- Archive systems; archive backup and recovery systems; archive media in use (transcription systems and policy);
- Access to archive via gateways, connection to high speed networks;
- Contents of the archives (EO missions, duration, coverage and volume), data curation aspects and preparation to long term preservation.

## 2.4 Questionnaire results vs. objectives

During the first period, the response rate reached the 10% of the 607 contacted organisations with the 5,6% of responses detailing data repositories and infrastructures, which are presented in the present document.

During the second period, more or less 50 additional answers (creations and modifications) were recorded leading to 120 out of 607 answers, approximately 20%.

The collected information is integrated with the provided reference documentation and the information available in the relevant web sites.

The rate is poor with regard to the number of contacted organisations. This is mainly due to the fact that:

- The contacted persons are not the ones who have the information, and they do not forward the questionnaire;
- They already have their own Website and dissemination system and have no time to answer;
- Although they have received the link to GENESI-DR Website, they do not feel concerned by a possible integration within GENESI-DR, or they do not want to share their data or the fact of sharing their data will necessitate too much effort or cost;
- They do not want to describe the kind of standards used, their data policy, the kinds of managed data, which they consider as internal information;

- They only provide partial information, since they do not really cope with the topics of data curation, used standards, sustainability or they follow different strategies according to the missions, which would necessitate too much time to explain;
- GENESI-DR portal was still at its early stages during the two periods, which might have led the contacted people to feel not very much concerned with the questionnaire objectives, waiting for the operational version.

# 3. Infrastructures

## 3.1 Introduction

The following considerations deal not only with infrastructures themselves but also with inter-operability. Interoperability is the ability of (EO, Science) systems to work together. One basic precondition to inter operate are infrastructures - backbone for the information traffic. But interoperations need more - rules and standards for the exchange and practical use for development of the standards.

The first part of this section deals with projects which intercede on behalf of interoperability and standardisation in the frame of science data. Considered projects contribute to needed standards and/or make precursor implementation of these. Several projects work as pathfinder and give valuable input to GENESI-DR. Several infrastructure initiatives and projects like HMA and INSPIRE are also taken into consideration in other GENESI-DR activities. But the approach is different. So e.g. in service activity 2 (SA2, see [17]) these initiatives serve as an input for the architectural design. In the frame of this document these initiatives serve as the environment of GENESI-DR. They are object of current and future cooperation and alignment. The investigated projects use backbones to make information exchange between science centers and cooperation.

The second part of this section summarizes these backbones and presents the Web-based collected information about existing European and national infrastructures.

## 3.2 Interoperability and standardisation efforts in Earth Sciences

One of the most fundamental challenges facing humanity at the beginning of the 21st century is to respond effectively to the global changes that are putting increasing pressure on the environment and on human society. Global changes, environmental degradation and sustainability are main aspects that need to be understood and managed. The demand for access to continuously increasing quantities of heterogeneous data and resources (mechanisms, infrastructure, tools, etc.) needed to achieve such objective is growing.

The need to face this increasing demand highlights two basic aspects on which the scientific community should focus:

- Data, coming from heterogeneous sources as needed in the different domains;
- Resources, to be used to exploit such data in an effective way.

Unfortunately, locating and accessing the right data, products, resources and other information needed to turn data into knowledge is not always easy.

Today, indeed, information about the state of the Earth, relevant services, data availability, project results and applications are accessible only in a very scattered way through different operators, scientific institutes, service companies, data catalogues, etc. Referring to remote sensing missions, only a limited community with specific knowledge of what to search for, is today in a position to collect, compile and thus exploit the necessary Earth Observation (EO) information.

It is as well not to be forgotten that each Earth Science community domain may need specific methods, approaches and working practices for gathering, storing and exchanging data and information.

This demand of the ES community shows the need for an efficient data infrastructure able to provide reliable, easy, long-term access to Earth Science data via the Internet, so to allow

Earth scientists and users to easily and quickly derive objective information and share knowledge based on all environmentally sensitive domains.

Interoperability appears to be a key issue in the development of an efficient and value adding data infrastructure.

### 3.2.1. Projects survey

This section provides a selection of projects/initiatives that could be considered as a response to the ES community needs.

In particular, their position with respect to interoperability and standardisation is evaluated.

| Project Name | Title and Description | Link | Interoperability and standardization key issues |
|---|---|---|---|
| CASPAR | Cultural, Artistic and Scientific knowledge for Preservation, Access and Retrieval, intends to provide tools and techniques for secure, reliable and cost-effective preservation of digitally encoded information for the indefinite future. | www.casparpreserves.eu | The interest in this project is driven by the fact that CASPAR methodological and technological solution:<br>• Is compliant to the OAIS Reference Model (the main standard of reference in digital preservation);<br>• Is technology-neutral (the preservation environment could be implemented using any kind of emerging technology);<br>• Adopts a distributed, asynchronous, loosely coupled architecture and each key component is self-contained and portable (it may be deployed without dependencies on different platform and framework);<br>• Is domain independent (it could be applied with low additional effort to multiple domains/contexts);<br>• Preserves knowledge and intelligibility, not just the “bits”;<br>• Guarantees the integrity and identity of the information preserved as well as the protection of digital rights. |
| Climate-G | Climate-G is a distributed testbed for climate change addressing challenging data and metadata management issues at a very large scale. | http://grelc.unile.it:8080/ClimateG-DDC | Climate-G implements OGC services. To enable geographical data sharing, search and discovery activities it adopts a distributed CMCC (Centro Euro-Mediterraneo per i Cambiamenti Climatici) metadata solution leveraging P2P and grid technologies, the GRelC Data Access and Integration Service. |

| Project Name | Title and Description | Link | Interoperability and standardization key issues |
|---|---|---|---|
| DEGREE | Dissemination and Exploitation of GRids in Earth science, aims to build a bridge linking the ES and GRID communities throughout Europe. An ES applications panel with a range of candidate applications suitable for porting to GRID will make sure key ES requirements for porting and deployment on the GRID middleware are identified, communicated and discussed within the GRID community. | http://www.eu-degree.eu/ | Its Work Package 2 deals with data management will perform a review of mostly used tools (database products), systems and standards for ES metadata in the ES community and will also evaluate and review advanced grid technology components to securely distribute, share and access ES data in an operation oriented environment. |
| DRIVERII | Digital Repository Infrastructure Vision for European Research, establishes a network of relevant experts and Open Access repositories. | www.driver-support.eu | It is worth remarking the attention paid by this project to interoperability.<br>DRIVER caters for the larger European digital library community through metadata and interoperability guidelines for exposing resources through repositories. The objective of these guidelines is to offer a best practice tool for the repository community, and to streamline repository developments across Europe.<br>More specifically, the “DRIVER Guidelines for Content Providers” aim at:<br>• Reaching interoperability on both syntactical (use of OAI-PMH and OAI-DC) and semantic (Use of vocabularies) layers;<br>• Providing orientation for managers of new repositories to define their local data-management policies;<br>• Directing managers of existing repositories to take steps towards improved services;<br>• Supporting developers of repository platforms to add supportive functionalities in future versions. |

| Project Name | Title and Description | Link | Interoperability and standardization key issues |
|---|---|---|---|
| D4Science | Distributed Collaboratories Infrastructure on Grid Enabled Technology for Science, aims to continue the path that the GÉANT, EGEE, and DILIGENT projects have initiated towards establishing networking, grid-based, and data-centric e-Infrastructures that accelerate multidisciplinary research. | www.d4science.eu | It focuses on data heterogeneity, sustainability and scalability.<br>Aiming to build a data-centric infrastructure, this project meets most of the ES identified requirements and is of particular interest for interoperability purposes because of the focus on heterogeneity, sustainability and scalability.<br>Thanks to the adoption of standards D4Science is playing a mediator and integrator role by aggregating and giving access to many heterogeneous digital repositories. |
| EMODNET | European Marine Observation and Data Network, aims to facilitate access to coherent data sets, to permit the recognition of data gaps and to shape a data collection and monitoring infrastructure directly suited to multiple applications. | http://ec.europa.eu/maritimeaffairs/emodnet_en.html | The design principles of EMODNET (formulated by the European Commission together with a specially-constituted Expert Group) address interoperability on the one hand, as well as data curation on the other hand. The project is developing standards across disciplines as well as within them and accompany data with statements on ownership, accuracy and precision |
| enviroGRIDS | Building Capacity for a Black Sea Catchment Observation and Assessment System supporting Sustainable Development: The Black Sea Catchment known as one of ecologically unsustainable development and inadequate resource management, which has led to severe environmental, social and economic problems. The EnviroGRIDS @ Black Sea Catchment project addresses these issues by bringing several emerging information technologies that are revolutionizing the way we are able to observe our planet. | http://www.envirogrids.net/ | A work package is dedicated to spatial data infrastructure (SDI) with particular focus on the specification of interoperability standards to be used in the project as well as on the definition of a Gridded Spatial Data Infrastructure with gridified hydrological models |

| Project Name | Title and Description | Link | Interoperability and standardization key issues |
|---|---|---|---|
| GENESI-DR | Ground European Network for Earth Science Interoperations - Digital Repositories, will allow scientists from different Earth Science disciplines located across Europe to locate, access, combine and integrate Earth-related data from space, airborne and in-situ sensors archived in large distributed repositories. | http://www.genesi-dr.eu/ | Its main characteristics are:<br>• Web service-oriented architecture (OpenSearch);<br>• Use of standards for metadata schema (RDF), use of DC vocabulary,<br>• Selection of two data curation metadata models: SAFE manifest file and PREMIS metadata dictionary.<br>See section 3.2.3. |
| GEOLAND2 | Integrated GMES Project on land cover and vegetation, aims to develop and demonstrate a range of reliable, affordable and cost efficient European geo-information services, supporting the implementation of European directives and their national implementation, as well as European and International policies. | http://www.gmes-geoland.info/index.php | It has the capability to handle heterogeneous data (both in situ and Earth Observation) although the purpose is limited to the derivation of geo-information. |
| GISHEO | On Demand Grid Services for Higher Education and<br>Training in Earth Observation: one of the technical objectives is to set-up and organize a virtual organization (VO) based on Grid technology for education, training and knowledge dissemination for Earth observation (EO) and related activities. | http://gisheo.info.uvt.ro | It focuses on education and training and Grid technology.<br>The drawback is the limited scope of involved data (only Earth Observation). |
| GMOSAIC | GMES services for Management of Operations, Situation Awareness and Intelligence for regional Crises, will provide the European Union with intelligence data that can be applied to early warning and crisis prevention as well as to crisis management and rapid interventions in hot spots around the world. | http://www.gmes-gmosaic.eu/home.html | It will study architectural models and prototype suitable solutions for security services provision, addressing specific requirements related to confidentiality of sensitive data and information handling. It will also provide recommendations for future sensors for security services. |

| Project Name | Title and Description | Link | Interoperability and standardization key issues |
|---|---|---|---|
| GSC DA | GMES Space Component Data Access, focuses on coordinating the access to space-based observation data to support GMES services. | http://cordis.europa.eu/icadc/fetch?CALLER=FP7_PROJ_EN&ACTION=D&DOC=187&CAT=PROJ&QUERY=011aa1a05777:7c91:569ece09&RCN=86262 | It provides a comprehensive and coordinated access to space data linking the different Earth Observation (EO) Data Providers and the different Service Providers by using coordinating functions. |
| HIDDRA | Highly Independent Data Distribution and Retrieval Architecture for Earth Observation Missions: proposes solutions to distribute data to the scientific community by means a novel architecture that solves the problem of the heterogeneity of data, networks and architectures. | http://www.springerlink.com/content/p4u544483np5304k/ | Data distribution is a common problem in observation missions. Data gathered by the satellite have to be distributed among different operational centres using different policies: as fast as possible, creating redundant copies, informing about the new available products, etc. To accomplish all these policies, the HIDDRA architecture takes into account features such as:<br>• Performance;<br>• Scalability;<br>• Availability of the system at any moment;<br>• Automatic retrieval technologies based on subscriptions;<br>• Flexibility, by dealing transparently with heterogeneous protocols;<br>• Data redundancy to increase the availability and to make information reliable. |

| Project Name | Title and Description | Link | Interoperability and standardization key issues |
|---|---|---|---|
| HMA | Heterogeneous Missions Accessibility: its objective is to define the interoperability concept across the ground segments of the European, Canadian and EUMETSAT missions which shall contribute to the initial phase of GMES. | http://wiki.services.eoportal.org/tiki-index.php?page=HMA%20Wiki | It focuses on interoperability of existing infrastructures and strong emphasis on metadata management (following OGC/ISO specifications and standards).<br>One of the core services is a catalogue that provides interactive access to query the availability of EO data from a large number of providers, all involved missions need to provide harmonized catalogue service interfaces (following OGC/ISO specifications and standards) that can be harvested and bundled by an entry point catalogue. |
| INSEA | Data Integration System for Eutrophication, is focused on the development of integrated management tools for coastal eutrophication assessment combining Models, Satellite Remote Sensing and in situ Measurements. | http://www.insea.info/ | One of the project deliverable, the “Metadata handbook”, provides metadata formats and data quality procedure guidelines to be followed by the INSEA partners. |
| MACC | Modelling Atmospheric Composition and Climate, is continuing operation of and improving the systems developed during the EU funded project GEMS. | http://www.ecmwf.int/publications/cms/get/ecmwfnews/1243612423986 | It focuses on services, data analysis and modelling systems for monitoring the global distributions of atmospheric constituents important for climate, air quality and UV radiation, with a focus on Europe. |
| MERSEA | Marine EnviRonment and Security for the European Area, aims to develop a European system for operational monitoring and forecasting on global and regional scales of the ocean physics, biogeochemistry and ecosystems. | http://www.mersea.eu.org/ | MERSEA provides for product information management:<br>• Product catalogue and product standard ISO19115 description;<br>• Product search facility. |

| Project Name | Title and Description | Link | Interoperability and standardization key issues |
|---|---|---|---|
| METAFOR | Common Metadata for Climate Modelling Digital Repositories): one of the major goals of this project is to create an Information Model that is common for all stages of both production and the use of climate model data. | www.metaforclimate.eu | It uses of standards for metadata and has defined a Common Information Model (CIM) with a controlled vocabulary.<br>The drawback is the limited scope of this project, which is specific of the environment ES sub-community. |
| MyOcean | MyOcean is the implementation project of the GMES Marine Core Service, aiming at deploying the first concerted and integrated pan-European capacity for Ocean Monitoring and Forecasting. It provides the best information available on the Ocean for the large scale (worldwide coverage) and regional scales (European seas), based on the combination of space and in situ observations, and their assimilation into 3D simulation models. | http://www.myocean.eu.org/index.php/project | Its main objective is the capability of combining and deriving information from heterogeneous data types (space and in situ) the ocean disciplines. |
| OASIS | OASIS project (Optimising Access to SPOT Infrastructure for Science) aims at optimising the access to the SPOT infrastructure to facilitate European space scientific research. | http://medias.obs-mip.fr/oasis | Products are free of charge for the European scientific communities who must satisfy certain specific conditions.<br>SPOT data will be downloaded by ftp. |

| Project Name | Title and Description | Link | Interoperability and standardization key issues |
|---|---|---|---|
| OMII-Europe | Open Middleware Infrastructure Institute for Europe, has been established to provide key software components for building e-Infrastructures within the European Research Area (ERA). | www.omii-europe.org | The project combines the interoperability concept with the Grid computing needs, since it aims to produce interoperable Grid components.<br>This would mean, for example, that a user can submit the identical job to any basic job execution service regardless of the Grid middleware platform it is running on.<br>In the short term, this will help other projects that rely on different platforms and make Grid services more accessible to researchers. In the long term, the work of OMII-Europe contributes to the ongoing global efforts to forward the state-of-the-art of Grid technology. |
| PARSE.Insight | Permanent Access to the Records of Science in Europe, aims to define roadmap & recommendations for developing the European e-Infrastructure enabled for maintaining the long-term accessibility and usability of scientific digital information in Europe. | http://www.parse-insight.eu. | PARSE.Insight has a key role among the presented projects as it aims to highlight the longevity and vulnerability of digital research data concentrating on the parts of the infrastructure needed to support persistence. Focussing on the Earth Science community and its requirements about accessibility to historical data, it considers the infrastructure needed to support persistence and understand-ability of these key assets over the long term, via reference to the OAIS reference model (ISO 14721).<br>Sustainability and standardisation is specifically addressed in its Work Package 6. |
| SAFER | Seismic eArly warning For EuRope, aims to fully exploit the possibilities offered by the real-time analysis of the signals coming from seismic networks for a wide range of actions, performed in a time interval of a few seconds to some tens of minutes. | http://www.saferproject.net | It focuses on accurate and fast algorithms, and real-time control. |

| Project Name | Title and Description | Link | Interoperability and standardization key issues |
|---|---|---|---|
| SeaDataNet | Pan-European infrastructure for Ocean and Marine Data Management provides integrated on-line access to a very comprehensive sets of multidisciplinary in-situ and remote sensing marine data, meta-data and products. | http://www.seadatanet.org/ | Although only focused on marine/ocean data, SeaDataNet is particularly interesting for the effort toward interoperability, which is the key to distributed data management system success. It is achieved in SeaDataNet by:<br>• Using common vocabularies;<br>• Adopting the ISO 19115 metadata standard for all metadata directories, resulting in harmonised metadata formats ;<br>• Using harmonised Data Transport Formats for data sets delivery;<br>• Using common quality control protocols and flag scale;<br>• Providing standard software tools;<br>• Providing a Central User Register and single-sign-on Authentication and Authorisation Services ;<br>• Using SOAP Web Services in the SeaDataNet architecture. |
| SOSI | SOSI is a project for developing innovative "Spatial Observation Services and Infrastructure" within the context of land monitoring initiatives at European and Member State levels. | http://wiki.services.eoportal.org/tiki-slideshow.php?page=SOSI%20Wiki&slide=1 | SOSI offers a distributed node-based infrastructure of Web-services following Service Oriented Architecture (SOA) principles and standards thus establishing access to a number of content services and one land cover generation processing service operated by the participating organizations.<br>Thanks to the attention to standardisation, the project supports interoperability, although the scope is limited to the environment ES discipline. |

**Table 1 - Infrastructures survey summary**

### 3.2.2. Earth Science needs in terms of infrastructures

As resulting from the analysis of the difficulties the Earth scientists are facing, the solution appears to be a dedicated global data infrastructure that supports efficient data management, high-performance processing and e-collaboration, which also helps community building.

A common dedicated data infrastructure providing access to historical data holdings, networks of sensors, broadband communications via ground and space, efficient, effective and distributed computing and storage resources to take care of and handle scientific tools, methodologies, data, etc. would support the Earth Science communities in deriving objective information and to *share knowledge regarding all environmental sensitive domains over a continuum of time* covering the past, the present and the future (considering historical measurements, real time assessments as well as short and long term predictions) and across a variety of *geographical scales* (from global scale down to very local facts).

Given the above, the following capabilities shall be supported as a priority by such an infrastructure:

- Capability, for Earth Science scientists, to discover and access distributed data from different data providers through a unique interface in a transparent and standardised way;
- Capability to provide fast access to large volumes of coherently maintained distributed data in an effective and timely way;
- Capability to access high performance computing resources through standardized interfaces;
- Capability, for data (archive or Digital Repository – DR) owners, to easily make available their data to a significantly increased audience with no need to duplicate them in a different storage system;
- Capability to ensure data curation and long term preservation of all types of Earth Science data.

As resulting from these high level requirements, **interoperability** appears to be a key issue in the development of an efficient and value adding data infrastructure.

Compliance to widely used **standards** (opportunely enhanced as needed) for **data management** as well as for processing services provision, appears to be the response to the interoperability requirement.

Different aspects need to be considered: data cataloguing, data discovery, digital right management for data, data access, metadata schemas, metadata exchange, ontology and data curation and preservation issues.

Standards for ES are defined in the framework of European and worldwide programs, projects, directives and initiatives such as SEIS, INSPIRE, GMES, GEOSS and OGC.

Important ES standards for data management are ISO15836:2003, ISO19115, ISO19119, ISO19139, OGC CS-W 2.0, FGDC and the Dublin Core.

The above listed high level user requirements, when analysed from a more technical point of view, can be turned into the below listed basic data infrastructure requirements:

- The data infrastructure shall establish and operate a network of at least the following services:
    - Discovery services: search for data sets and data services on the basis of the content of the corresponding metadata and display of the content of the metadata;

  - View services: making it possible, as a minimum, to display, navigate, data sets and any relevant content of metadata;
  - Download services: enabling copies of complete spatial datasets, or of parts of such sets, to be downloaded;
  - Transformation services: enabling spatial data sets to be transformed;
- The data infrastructure shall expose Web service based interfaces (e.g. on the basis of the OGC service definition) for discovering and accessing heterogeneous and federated data and services resources;
- The different metadata schemas supported by the data infrastructure shall be harmonised towards a common standard - one for data and application management;
- User/role based access control to metadata and data has to be supported via standardized authentication and authorization mechanisms;
- The data infrastructure shall support the different data formats *which are widely used in the ES community, such as* NetCDF and HDF file;
- The data infrastructure shall support for several and efficient data transfer technologies, such as P2P or other commonly used by ES *access protocols, such as* OpenGIS and OPenDAP networks access protocols;
- The data infrastructure shall s*upport for special types of databases* beyond the relational model, such as spatial databases, which represent a standard backend for the OpenGIS web-map services.
- The data infrastructure shall create metadata for all spatial datasets and services, and keep these up to date following the INSPIRE directives;
- The data infrastructure shall allow for access to high performance computing resources. Grid infrastructure and systems respond to this requirement, since Grid computing permits the sharing of resources between institutions and allows for scaling up computing power and storage capacity.
- In-depth examination of the subject can be found in “Grid Computing for Earth Science”, Eos, Vol. 90, No. 14, 7 April 2009 by P. Renard, V. Badoux, M. Petitdidier, and R. Cossu.
- An interface standard for Grid technology towards OGC services is defined within **OGF** (**Web Processing Service, WPS**). The ultimate objective is to easily interface the ES environments with the Grid environments.
- The data infrastructure shall ensure compliance to a standard data curation model (such as OAIS).

For some of these requirements ad-hoc solutions are available or investigated as shown by the survey presented in the previous section, but these cannot be integrated as such into new infrastructures.

### 3.2.3. How interoperability is addressed by GENESI-DR

GENESI-DR (Ground European Network for Earth Science Interoperations - Digital Repositories, http://www.genesi-dr.eu/) is an ESA-led, European Commission (EC)-funded two-year project, aiming at providing reliable, easy, long-term access to Earth Science data. GENESI-DR allows scientists from different Earth Science disciplines located across Europe to locate, access, combine and integrate Earth-related data from space, airborne and in-situ sensors archived in large distributed repositories. A dedicated infrastructure providing transparent access to all these data supports Earth Science communities by allowing them

to easily and quickly derive objective information and share knowledge based on all environmentally sensitive domains.

GENESI-DR is a response to most of the Earth scientists needs for interoperability.

The Central Discovery Service allows to query information about data existing in heterogeneous catalogues, at federated DR sites in a transparent way and can be accessed by users via web interface, the GENESI-DR Web Portal, or by external applications via open standardized interfaces (OpenSearch-based) exposed by the system.

Attention to standardisation in modelling information, flexibility and scalability of the architecture allow easy integration of new Digital Repositories.

GENESI-DR information objects follow the ISO 19115 standard for describing geographic information and services (and the corresponding XML schema implementation ISO 19139) and the INSPIRE Implementing Rules to use ISO 15836 (Dublin Core). Together they provide a profile that fits very closely with INSPIRE guidelines but allows more precision and models metadata about geographic information where the usage of Dublin Core is maximized.

To share information about datasets and collection available on the Digital Repository nodes the GENESI-DR information model is defined as an RDF model rendered in XML.

Different and efficient data transfer technologies such as HTTPS, GridFTP and BitTorrent are used to guarantee easiness and fast access to large volumes of distributed data;

GENESI- DR foresees services to enable expert users exploiting computational and network resources in order to produce the final desired product. This means either passing input data to a processing service (an OGC Web Processing Service, for instance) available at some site or running a user application/algorithm on Grid resources on specified data sets.

Another objective of GENESI-DR is the adoption of a data curation and preservation strategy in order to preserve the historical archives and the ability to access the derived user information as both software and hardware transformations occur.

In this context the OAIS model (information model), the PREMIS vocabulary and SAFE manifest file have been analysed for the definition of data preservation metadata.

GENESI-DR – being a data centric and interoperability based e-infrastructure - can represent the most comprehensive solution to the ES needs: interoperability, data heterogeneity management and multi-disciplinarity requirements are in fact met and considered in the appropriate direction.

At the same time inside this framework enhancements are desirable in order to support and satisfy the dynamically emerging ES needs to maximize interoperability, operativity and inter-disciplinary collaboration.

## 3.3 European Infrastructures

### 3.3.1. Introduction

The following infrastructures are characterized by a multi-project purpose and high investments and maintenance costs not affordable for a single project. But they represent the backbone for scientific exchange and collaboration. The following survey gives the possibility for future cooperation.

The information which has been relevant to collect about these infrastructures is:

- Their name and associated Website;
- The background of their creation and their objectives;

- Their network characteristics (speed, specific used technologies, …);
- Their main services and specific used technologies.

### 3.3.2. EGEE

EGEE (Enabling Grids for E-sciencE) (see Website: http://www.eu-egee.org/) is a very large multi-disciplinary grid infrastructure, which brings together more than 140 institutions to produce a reliable and scalable computing resource available to the European and global research community. At present, it consists of approximately 300 sites in 50 countries and gives its 10,000 users access to 80,000 CPU cores around-the-clock.

EGEE-III, co-funded by the European Commission, aims to expand and optimise the Grid infrastructure, which currently processes up to 300, 000 jobs per day from scientific domains ranging from biomedicine to fusion science.

The Production Service infrastructure is a large multi-science Grid infrastructure, federating some 250 resource centres world-wide, providing some 40.000 CPUs and several Petabytes of storage. This infrastructure is used on a daily basis by several thousands of scientists federated in over 200 Virtual Organizations on a daily basis. This is a stable, well-supported infrastructure, running the latest released versions of the gLite middleware.

EGEE supports applications from many scientific domains, such as astrophysics, biomedicine, computational chemistry, earth sciences, high energy physics, finance, fusion, geophysics and multimedia. In addition, there are several applications from business sectors running on the EGEE Grid, such as applications from geophysics and the plastics industry.

EGEE offers a portal and single access point for support to end-users, with security and policy principles, based on authentication and an agreed network of trust.

### 3.3.3. ESA G-POD

The ESA G-POD (Grid-Processing On Demand) (see Website: http://eogrid.esrin.esa.int) is a grid based operational infrastructure to provide improved data and tools access to science users, thus respecting the constraints imposed by the ESA data policy.

The ESA G-POD offers approved users with the possibility to perform bulk processing and/or validation of their own algorithms exploiting the large ESA earth-observation archive together with ESA available grid computing and dynamic storage resources.

The ESA G-POD portal integrates several technologies and distributed services to provide and end-to-end application process capable of being driven by the end-user. The portal integrates:

- User authentication services;
- Web Mapping services for map image retrieval and data geo-location;
- Access to metadata;
- Catalogues such as those contained in MUIS (ESA Multi-mission User Interface System) to identify the datasets of interest and access the AMS (ESA Archive Management System) archive to retrieve the data;
- Access to grid FTP transfer protocols to stage the data to the grid;
- Access to the grid Computing Elements (more than 200) and Storage Elements (more than 120 TByte) to process the data and retrieve the results.

### 3.3.4. EUFAR

EUFAR (European Fleet for Airborne Research) (see Website: http://www.eufar.net/, *Transnational Access Project)* is an Integrated Infrastructure Initiative of the Sixth Framework Program of the European Commission. EUFAR brings together 24 leading European institutions and companies involved in airborne research, operating 24 instrumented aircraft.

EUFAR aims at integrating the activities of the European operators of airborne research infrastructures, through:

- Co-ordinating the network for exchanging knowledge, sharing developments, and building the unified structure that is required for improving access to the infrastructures;
- Providing users with Transnational Access (TA) to the infrastructures;
- Extending TA to national funding sources;
- Promoting airborne research in the academic community;
- Developing research activities in airborne instrumentation.

The applicant must satisfy some Eligibility Criteria:

- Come from a member or associate country of the European Union;
- Come from a different European Country than the one operating the aircraft;
- Have no personal/direct links with the selected aircraft operator;
- Have the resources to ensure that the generated data will be analysed and disseminated;
- Be prepared to write a description of work.

The selection of applicants will be based on:

- Logistic & Feasibility;
- Scientific merit;
- Educational Impact.

Different kinds of experiments and measurements are produced and classified as follows: ocean, wind – turbulence, radiation, pressure, temperature, geophysics, humidity – ice, clouds microphysics and aerosols. As well, different kinds of calibration facilities are available: calibration Fast FSSP, humidity calibration, radiance, temperature calibration.

### 3.3.5. GÉANT

The GÉANT project (see Website: www.geant.net/) started from a collaboration between 26 National Research and Education Networks (NRENs) across Europe, the European Commission, and DANTE (co-funded by The European Commission within its 5$^{th}$ R&D Framework Programme). DANTE was the project's co-ordinating partner. The project began in November 2000 and was originally due to finish in October 2004; however, because of the project's success, and in order to permit a smooth transition to the next generation of the network (GÉANT2), the project had been extended until 30 June 2005.

The GÉANT network is a pan-European backbone which connects Europe's national research and education networks (NREN). The GÉANT network topology is continually upgraded to provide the best possible service to the NRENs, and thereby ultimately to Europe's research and education community.

GÉANT applies a distributed access methodology (European Distributed Access) which provides a number of access points for connection with other world regions. The existence of the multi-gigabit core enables this connectivity to be efficiently and equitably shared among the connected countries. The multi-location approach facilitates the most cost-effective provision of connectivity.

GÉANT now has a total of 12Gbps connectivity to North America, 2.5Gbps to Japan and 155Mbps to South Africa. These links provide connections between GÉANT and its equivalent networks in North America (Abilene, CANARIE and ESnet), Japan (SINET) and South AFrica (TENET).

Other research networking projects have expanded GÉANT's connections to other world regions: the ALICE and EUMEDCONNECT initiatives both provide connectivity to GÉANT, for the Latin American and Mediterranean regions respectively. In future, connectivity to the Asia-Pacific region will be provided through the TEIN2 project.

GÉANT has developed what is now called the European research networking model, a hierarchical model which is now exported and used for developing other networks.

There has been a second version: GEANT2 network (co-funded by the European Commission in its 6th R&D Programme), see Figure 2 below, which connects 34 countries through 30 national research and education networks (NRENs). GEANT2 is also operated by DANTE. It can be considered as the first international hybrid research and education network, based on dark fibre for greater network performance.

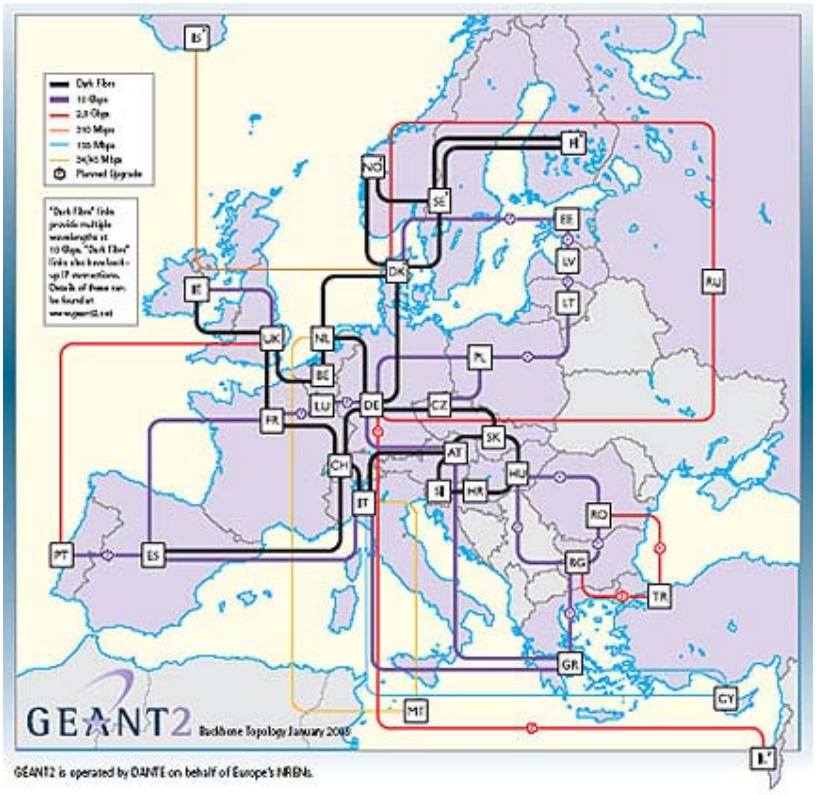


**Figure 2 - GEANT2 topology (connectivity between the NREN)**

It has to be noted that even a project of advanced telemedicine system links four European hospitals using GÉANT and local research networks, thanks to its speed performance and high connectivity.

## 3.3.6. GMES

### 3.3.6.1. Overview

Global Monitoring for Environment and Security (GMES) has been established to fulfil the growing need amongst European policy-makers to access accurate and timely information services to better manage the environment, understand and mitigate the effects of climate change and ensure civil security.

Under the leadership of the European Commission, GMES relies largely on data from satellites observing the Earth. Hence, ESA – in accordance with the European Space Policy – is developing and managing the Space Component for the initiative. The European Commission, acting on behalf of the European Union, is responsible for the overall initiative, setting requirements and managing the services.

To ensure the operational provision of Earth-observation data, the Space Component includes a series of five space missions called 'Sentinels', which are being developed by ESA specifically for GMES. In addition, data from satellites that are already in orbit, or are planned will also be used for the initiative. These 'Contributing Missions' include both existing and new satellites, whether owned and operated at European level by the EU, ESA, EUMETSAT and their Member States, or on a national basis. They also include data acquired from non-European partners. The Space Component forms the European contribution to the worldwide Global Earth Observation System of Systems (GEOSS).

The acquisition of reliable information and the provision of services form the backbone of Europe's GMES initiative. Services will be based on data from a host of existing and planned Earth observation satellites from European and national missions, as well as a wealth of measurements taken in situ from instruments carried on aircraft, floating in the oceans or positioned on the ground.

The GMES in situ component is based on an observation infrastructure owned and operated by the large number of national and European stakeholders coordinated, in some cases, within the framework of European and international networks. The in situ component is under the coordination of the European Environment Agency (EEA).

### 3.3.6.2. The GMES multi-mission Ground Segment

The GMES Ground Segment facilitates the harmonised access and distribution of Earth-observation data and products from all the dedicated Sentinel satellites and Contributing Missions for GMES Services.

The GMES multi-mission Ground Segment also comprises the mission control to operate the Sentinel satellites and the payload ground segment to handle data received from the dedicated Sentinel satellites and to elaborate the data into products for GMES Services.

The Ground Segment, which is spread out geographically, relies on existing national public and private facilities and infrastructure that are managed coherently.

Since the GMES Space Component includes a complex and evolving constellation of Contributing Missions, it is envisaged that each of these missions will provide its own ground segment services. ESA, several ESA Member States, EUMETSAT and other Third Party operators have developed, or are in the process of developing, various satellite programmes. While these programmes primarily fulfil their own objectives, they may also offer essential capacities for GMES.

As part of the GMES Ground Segment, the Coordinated Data Access System (CDS) will ensure the harmonised provision of data from the Sentinels as well as from the Contributing Missions. The CDS will also support the building of the Earth Observation product collections defined within the so-called Data Access Portfolio such as the European wall-to-wall multi-spectral mosaic to be used by the Land Service.

A dedicated GMES Space Component Data Access website has been set up for GMES Services to access Earth Observation data.

The GMES Ground Segment will be addressing the challenge to elaborate, distribute and archive, each day, around 10 000 Gigabytes of data received from the full constellation of Sentinel-1, Sentinel-2 and Sentinel-3 spacecrafts.

#### 3.3.6.3. The GMES Space Component (GSC) Data Access

The GMES Space Component (GSC) Data Access system is the interface for accessing the Earth Observation products from the GMES Space Component. The system overall space capacity relies on several EO missions contributing to GMES, and it is continuously evolving, with new missions becoming available along time and others ending and/or being replaced.

The GMES Space Component Data Access (GSCDA) provides a comprehensive and coordinated access to all GMES space data, allowing the capacity:

- To link transparently the different EO Data Providers and the various GMES Service Providers using coordinating functions;
- To create synergy and sustainability across the various contributing missions;
- To facilitate the data access for the services and aim at long-term data reliability beyond single missions.

GSCDA data and services are accessible in the form of DataSets, which are pre-defined collections of coherent multi-mission products responding to specific users needs. These DataSets have been derived from the requirements of the GMES Service Projects, and after trading-off with the overall system capacity.

The overall assortment of the available DataSets is called the GSC Data Access portfolio (DAP). The DAP undergoes evolution cycles based on formal reviews, where updated requirements and evolving space systems capacity are taken into account.

The DAP specifies the DataSets that are available from the GSCDA.

The GSC-Data Access project is managed by the European Space Agency as an integral part of the GMES Space Component Programme.

### 3.3.7. HiSEEN

HiSEEN, High-Speed ESA Earth Observation Network is an ESA funded operational VPN which allows high-volume data transfers to take place between Earth Observation facilities, using the GÉANT Europe's National Research and Education Network (NREN), at rates in the range of 20-200Mbps.

The front-end infrastructure of each centre is made up of an access router and of a VPN firewall. A full-mesh IP VPN built over the HiSEEN provides encrypted connectivity between the internal networks of the centres.

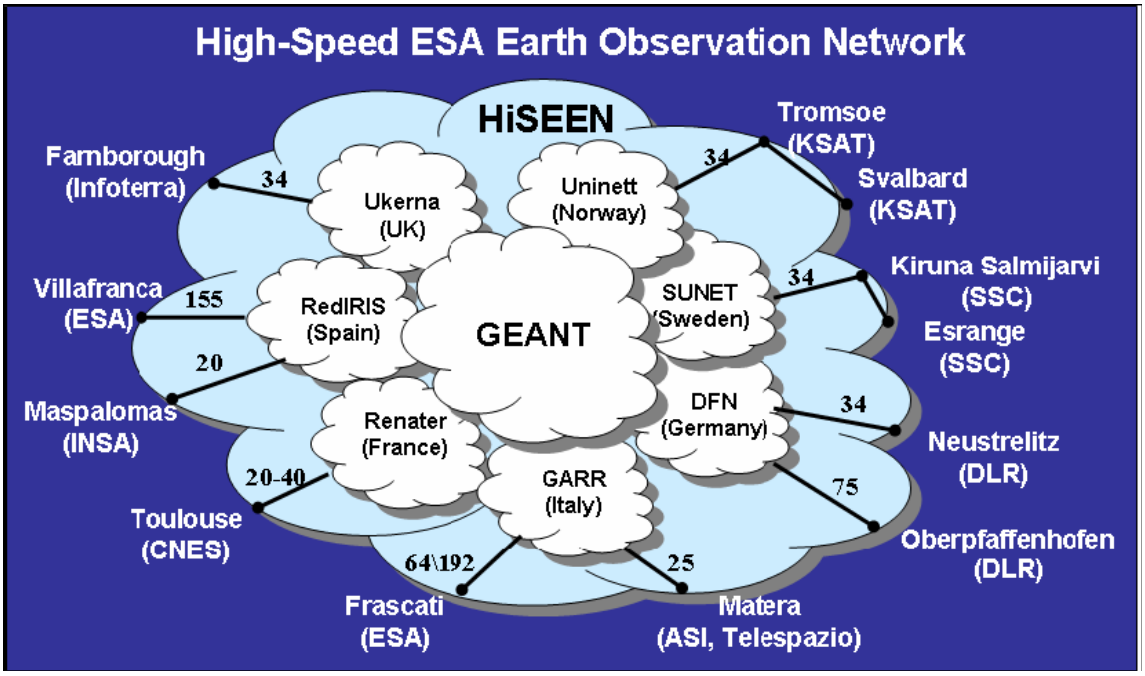


**Figure 3 - HiSEEN infrastructure**

HiSeen infrastructure is based upon ESA facilities spread among different partners'sites (as shown by Figure 4:

- in yellow, these are the NRT Processing Centres at ESRIN and KIRUNA;
- in green, these are the Processing and Archiving Centres (PAC) in different European locations (Italy, Sweden, Spain, France, United Kingdom, Germany).

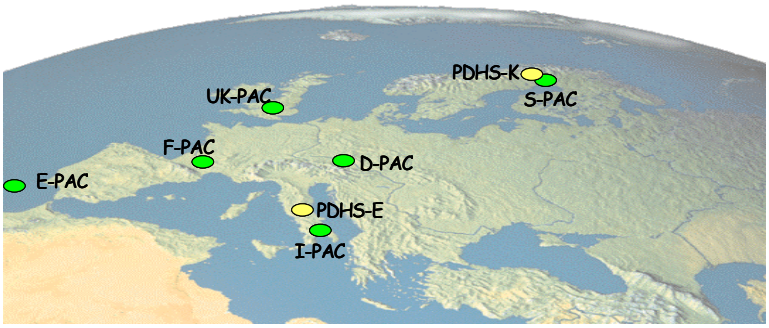


**Figure 4 - ESA PAC facilities**

## 3.4 National Infrastructures

### 3.4.1. Introduction

One of GENESI-DR objectives is to provide valuable information not only for universities but also for schools. Some very simple applications are planned to be designed to help people understand the world in which they live (for example: pollution map, vegetation evolution

map, floods study, etc.). These pre-defined products will be computed and distributed by the GENESI-DR infrastructure for these specific needs.

Another objective is to disseminate research/scientific data results and provide an easy access for scientists to Earth Science products for their studies and multidisciplinary applications.

Therefore it seems appropriate to have a list of the major European research and education networks, as possible contacts interested in both GENESI-DR objectives.

The national infrastructures presented in this section correspond to the European National Research and Education Networks (NREN). A NREN is a specialised internet service provider dedicated to supporting the needs of the research and education communities within a country.

It is usually distinguished by support for a high-speed backbone network, often offering dedicated channels for individual research projects. NRENs are usually the places where new Internet protocols are introduced before deployment within the Internet.

There are two major European organizations or associations dealing with NREN:

- TERENA (Trans-European Research and Education Networking Association) is an association of European NRENs: it offers a forum to collaborate, innovate and share knowledge in order to foster the development of Internet technology, infrastructure and services to be used by the research and education community (see Website at: http://www.terena.org/).
- DANTE (Delivery of Advanced Network Technology to Europe) - runs GÉANT2 backbone network on behalf of European NRENs. DANTE plans, builds and operates advanced networks for research and education (see Website: http://www.dante.net/). It is owned by European NRENs (national research and education networks), and works in partnership with them and in cooperation with the European Commission. DANTE provides the data communications infrastructure essential to the development of the global research community.

The following European NREN have been listed, and they are all accessible via GEANT2:

- ACOnet - Austrian NREN;
- AMRES - Serbian NREN;
- ARNES - Slovenian NREN;
- Belnet - Belgian NREN;
- CARNet - Croatian NREN;
- CESNET - Czech NREN;
- CYNET - Cypriot NREN;
- DFN - German NREN;
- EENet - Estonian NREN;
- FCCN - Portuguese NREN;
- Consortium GARR - Italian NREN;
- GRNET - Greek NREN;
- HEAnet - Irish NREN;
- HUNGARNET - Hungarian NREN;
- JANET (JANET(UK)) - United Kingdom NREN;

- JSCC - Russian NREN;
- SigmaNet - Latvian NREN;
- LITNET - Lithuanian NREN;
- MARNET - Macedonian NREN;
- NORDUnet - Nordic backbone network
    - FUNET - Finnish NREN;
    - RHnet - Icelandic NREN;
    - SUNET - Swedish NREN;
    - Forskningsnettet - Danish NREN;
    - UNINETT - Norwegian NREN;
- PIONIER (PCSS) - Polish NREN;
- RedIRIS - Spanish NREN;
- RENATER - French NREN;
- RESTENA - Luxembourg NREN;
- RNC / RoEduNet - Romanian NREN;
- SANET - Slovakian NREN;
- SURFnet - Dutch NREN;
- SWITCH - Swiss NREN;
- ULAKBIM - Turkish NREN;
- University of Malta CSC - Maltese NREN.

The information provided in this document comes from the NREN own Website as well as from Wikipedia. The main aspects which are presented are: the physical network, its capacities and specific used technologies, the links to other infrastructures or programmes.

### 3.4.2. Austria infrastructure: ACOnet

The Austrian Academic Computer Network (ACOnet) (see Website: http://www.aco.net/ is a national backbone network for science, research and education. ACOnet provides a powerful network-infrastructure based on DWDM-technology and Gigabit Ethernet for data communication between academic institutions, international networks and offers full Internet connectivity. ACOnet operates points of presence in the following locations: Vienna, Linz, Graz, Salzburg, Innsbruck, Klagenfurt, Eisenstadt, Leoben and Dornbirn.

Since autumn 2007 ACOnet is being upgraded to a wavelength- transparent fibre optic backbone. This allows multiple 10 Gbit transmission paths per route. To provide a stable and failsave network, all ACOnet points of presence will be connected in a redundant manner. Vienna, Linz and Graz have already been upgraded to the new technology. The completion of the whole fibre optic network is planned for the end of 2008, leading to the following features (see Figure 5):

- Gigabit- and 10-Gigabit Ethernet backbone;
- Multi-Gigabit Internet Access;
- 10-Gigabit connection to European and international research and education networks via GÉANT2;

- 20-Gigabit connection to the Vienna Internet eXchange for a quick handling of regional data traffic;
- Cross-border fibre optic lines to CESNET (.cz) and to SANET (.sk);
- Global IPv4/IPv6 connectivity;
- Global IPv4/IPv6 multicast-connectivity.

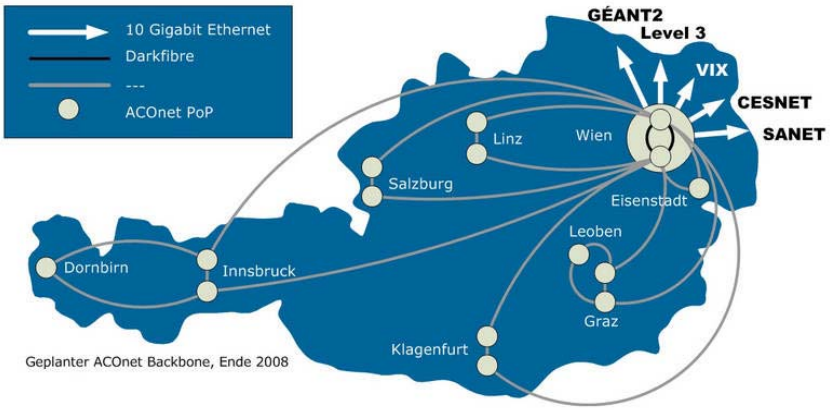


**Figure 5 - ACOnet network**

### 3.4.3. Czech Republic infrastructure: CESNET

CESNET (see Website: http://www.ces.net), association of legal entities, was held in 1996 by all universities of the Czech Republic and the Czech Academy of Sciences. Its main goals are:

- operation and development of the Czech NREN;
- research and development of advanced network technologies and applications;
- broadening of the public knowledge about the advanced networking topics.

The actual CESNET2 network is built around the DWDM backbone ring connecting Prague, Brno, Olomouc and Hradec Kralove, continuously expanding to further cities.

Remaining backbone lines of the CESNET2 network are based on the mixture of Ten-Gigabit and Gigabit Ethernet and Packet Over SONET (POS). The essential idea is to build the backbone network in small rings encompassing no more than four cities (points of presence). Such topology provides redundant network offering small hop counts and delays.

The network has following independent international connections (see Figure 6):

- 10 Gbps to GÉANT, used for academic traffic;
- 800 Mbps to Telia, used for commodity traffic;
- 10 Gbps to SANET, academic network of Slovakia;
- 10 Gbps to ACOnet, academic network of Austria;
- 10 Gbps to PIONIER, Polish optical academic network.

Network provides multicasting capability and is a part of the MBone network. Videoconferencing services based on the MBone technology are offered to the network users. IP version 6 is provided as a standard service.

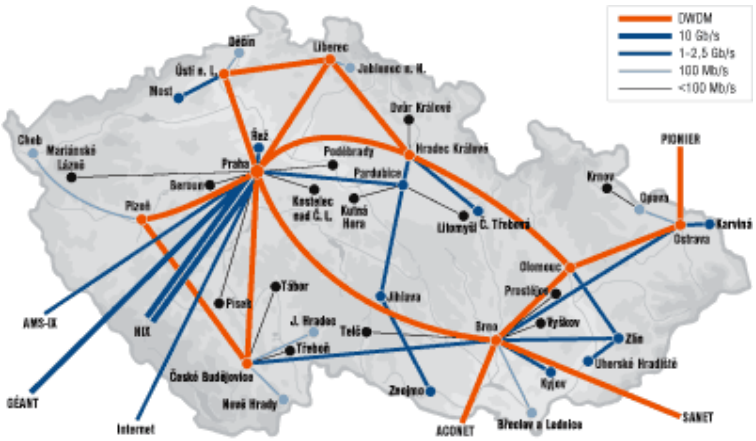


**Figure 6 - CESNET network**

### 3.4.4. Germany Infrastructure: DFN

Deutsches Forschungsnetz (DFN) ("German Research Network") is the German national research and education network (NREN) used for academic and research purposes. It is managed by the scientific community organized in the voluntary association "Verein zur Förderung eines Deutschen Forschungsnetzes e.V." which was founded in 1984 by universities, non-university research institution and research-oriented companies to stimulate computerized communication in Germany.

The current backbone X-WiN, operating on multi-gigabit speeds and linked with 10G links to GÉANT2, replaced its predecessor G-WiN in 2006 (see Figure 7).

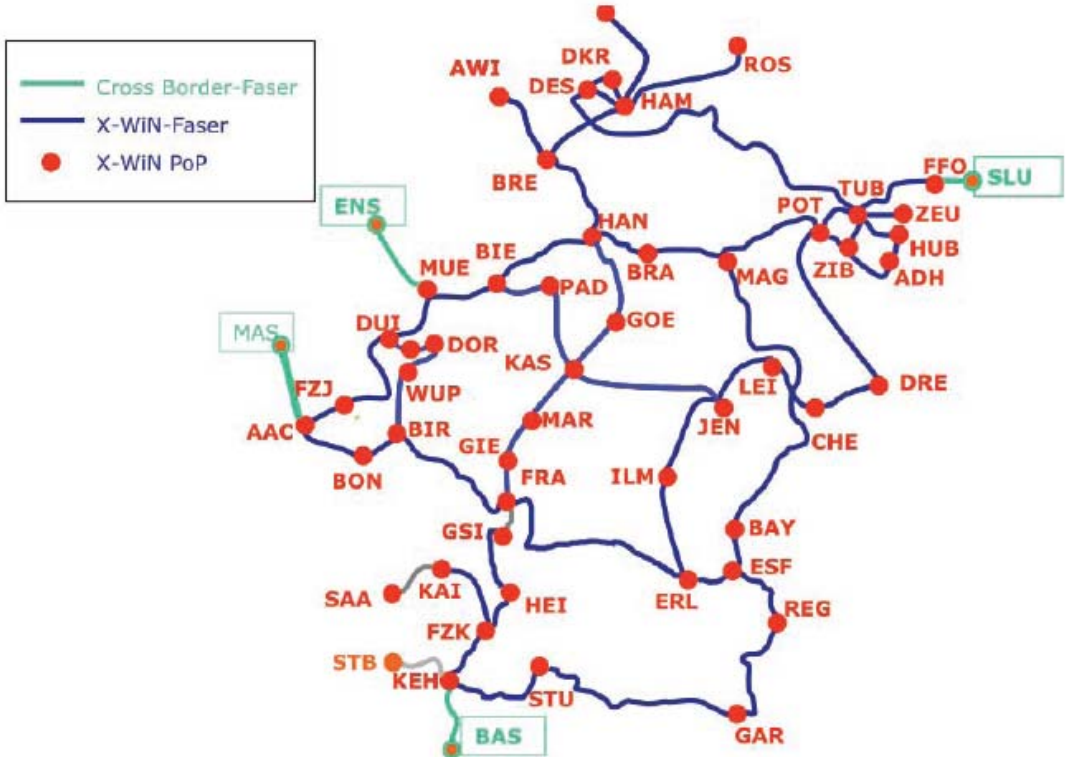


**Figure 7 - X-WiN Fibre based DFN infrastructure**

### 3.4.5. Italian Infrastructure: GARR

GARR ("Gestione Ampliamento Rete Ricerca") is the Italian Academic and Research Network.

The main institutional tasks of GARR towards its own Community is:

- to implement and to manage the networking interconnection service and the interconnection service to the other european and worldwide Research networks and to Internet in general;
- to supply operational and application networking services;
- to support the coordination and collaboration among the Research activities (national and international level) via telematic services, including the research and development in telematics itself;
- to support the dissemination, the information update and the exchange of knowledge in telematics, also organising Workshops and Courses.

ENEA, CNR and ASI partake to GARR (see Website: www.garr.it).


**Figure 8 - GARR infrastructure**

### 3.4.6. Greece Infrastructure: GRNET

GRNET (Greek Research and Technology Network) is the national research and education network of Greece. GRNET develops and provides advanced services of national and international internet access to the Research, Academic and Education communities of Greece, with its gigabit GRNET2 network and the Virtual NOC supportive scheme. See Website at: http://grnet.gr/.

GRNET's networking infrastructure serves as the foundation for advanced computing applications, taking advantage of the processing/storage clusters that reside in different parts of the country and sophisticated middleware developed by European collaborators. Additionally, GRNET is the coordinator of the SEE Federation for EGEE, the pan-European Grid infrastructure.

The network connects:

- 27 Universities;
- 15 Technical Universities;
- 33 Research Institutions;
- 12,673 schools to the Pan-European Research and Education Network, GEANT.

GRNET2 constitutes a new generation optical fiber of WavelengthDivisionMultiplexing – WDM technology at extra high speeds (1-2,5 Gbps). All the nodes are based on routers of Gigabit speeds and are interconnected with a network of 2.5 Gbps speeds over DWDM technology with leased wavelengths from the incumbent (OTE). See Figure 9.


**Figure 9 - GRNET infrastructure**

### 3.4.7. United Kingdom Infrastructure: JANET

JANET is a private British government-funded computer network dedicated to education and research. All further- and higher-education organisations in the UK are connected to JANET, as are all the Research Councils; the majority of these sites are connected via 20 metropolitan area networks or regional networks, emphasising that JANET connections are not just confined to a metropolitan area across the UK. The network also carries traffic between schools within the UK, although many of the schools' networks maintain their own general Internet connectivity. The name was originally a contraction of Joint Academic NETwork but it is now known as JANET in its own right.

It is linked to other European and worldwide NRENs through GEANT, has a private connection to its equivalent CERNET in China and peers extensively with other ISPs at Internet Exchange Points in the UK. Any other networks are reached via transit services from commercial ISPs.

JANET is operated by JANET(UK), formerly known as UKERNA (the United Kingdom Education and Research Networking Association), who are also responsible for the .ac.uk and .gov.uk domains. It is funded by JISC, the Joint Information Systems Committee. See Website at: http://www.ja.net.

They have implemented what is called JANET Roaming developed from the Location Independent Networking (LIN) concept for providing simple authenticated independent network access for visitors to JANET connected organisations. It comprises an infrastructure to enable guest users to use their own home network registered user credentials (eg. username@foo.ac.uk and home password) to gain authenticated independent network access at participating organisations, without any administrative burden or added complexities - both for the user and the local IT staff.

There is also JANET Aurora which is a dark-fibre network to support research on photonics and optical systems. It interconnects research groups at five Universities, and provides access to intermediate locations along each fibre path where additional equipment can be sited. This is a national facility funded via JISC. The characteristics are:

- Individual span length <= 50km;
- Chromatic dispersion per span <= +1100ps/nm;
- Polarisation Mode Dispersion (PMD) <= 2.5ps (between JANET access points);
- Optical loss per span <= 17dB.

The system “JANET Netsight” provides visibility of the performance of the JANET network. The system is available to JANET connected organisations and the regional network providers. It enables to monitor the performance of the whole of the JANET backbone, by polling all the other edge points on the backbone. The information collected is then made publicly available on the web on the Netsight system itself, in both tabular and graphical formats.


**Figure 10 - JANET infrastructure with its Regional Networks**

### 3.4.8. Irish Infrastructure: HEAnet

HEAnet (Ireland's National Education & Research Network), provides high-speed Internet access to academic institutions in Ireland direct to European and USA networks. The network is funded by the Higher Education Authority and is a WAN-based network. See Website at: http://www.heanet.ie/.

In 1984 the network was created with assistance from Dublin Institute of Technology, Dublin City University, University of Limerick, Trinity College, Dublin, University College, Cork,

University College, Dublin and University College, Galway. Today approximately 50 organisations support its work, including the Institutes of Technology as well as other colleges and non-education organisations in Ireland.

HEAnet helped set up INEX, the Internet Neutral EXchange, INEX provides a modern layer 2 ethernet switching network supporting IPv4, IPv6 and facilitating multicast peerings on several vlans, with supported access speeds of 10Mb/s, 100Mb/s, 1Gb/s and 10Gb/s.

HEAnet has been using leased dark fibre in Ireland since 2003. HEAnet currently leases approximately 3000Km of dark fibre in Ireland.

The aim of GRID Ireland is to provide a research platform for scientists allowing them to construct virtual organisations removing the boundaries of their home institutes. It operates in support of research projects such as Cosmogrid, which involves Grid computing across many scientific disciplines.

The infrastructure consists of 17 Grid gateways which connect sites (Universities, Colleges, Research Insitutions) to the Grid, allowing local users access to computing resources throughout Ireland and worldwide.

HEAnet also hosts a popular mirror service, ftp.heanet.ie, which acts as a mirror for projects such as SourceForge.net, Debian, Ubuntu and Red Hat.

### 3.4.9. Nordic Infrastructure: NORDUNET

NORDUnet is an international collaboration between the Nordic national computer networks for research and education. The members of NORDUnet are:

- SUNET of Sweden (Swedish Unversity Network) (www.sunet.se)
- UNINETT of Norway (Norwegian Research Network) (www.uninett.no)
- FUNET of Finland (Finnish University and Research Network) (www.funet.fi)
- Forskningsnettet of Denmark (Danish Research Network) (www.forskningsnettet.dk)
- RHnet of Iceland (Icelandic University Research Network) (www.rhnet.is)

NORDUnet interconnects these networks and connects them to the worldwide network for research and education and to the general purpose Internet. NORDUnet provides its services by a combination of leased lines and Internet services provided by other international operators.

NORDUnet currently operates a dual stack IPv4 and IPv6 production network. IPv6 development activities take place in a dedicated network running on top of NorthernLight infrastructure.

NorthernLight is a high-performance network facility, built to provide end-to-end network services to demanding users, to serve as a testbed for development of new networks and services, and to allow the Nordic research community to participate in international lambda networking activities.

The name "lambda networking" derives from the ideas of using different wavelengths (colours) of light on optical fibres, creating multiple channels on a single fibre pair. The vision is that international research groups, projects, or activities can be provided their own worldwide wavelength or "light path", enabling researchers to build dedicated, application-specific networks with large capacities. We call these networks Optical Private Networks (OPN).

### 3.4.10. Poland Infrastructure: PIONIER

The programme "PIONIER: Polish Optical Internet - Advanced Applications, Services and Technologies for the Information Society" has been prepared in 1999 (see Website:

http://www.pionier.gov.pl). In 2000 the project was accepted by the State Committee for Scientific Research as a continuation of former programs for building the Information Infrastructure for Science in Poland.

Research-deployment objectives of the PIONIER Program are the following:

- development and deployment of a new architecture for the Optical Internet;
- development and deployment of advanced services for the Information Society;
- development and deployment of a new economical model for the Optical Internet.

One of the main priorities of the PIONIER Program is to build a new generation optical network- the PIONIER network for short. The network topology includes over 6000 km of its own optical fibers and will connect 21 optical academic metropolitan area networks. The DWDM technology, up to 32 lambdas, is used starting with 10 Gb/s Ethernet.

The PIONIER Network (see Figure 11) has optical international links to research and education networks such neighboring countries as Germany, the Czech Republic (CESNET), Slovakia (SANET), Ukraine, Bielarus, Lithuania and Russia. For the benefits of the European co-operation it provides international connectivity to the pan-European network GÉANT through 10 Gb/s Ethernet and 10 Gb/s Packet over Sonet direct connections between the PIONIER Network and GÉANT PoP in Poznan. Moreover, two 2.5 Gb/s links provide general Internet access for the whole academic community in Poland. The PIONIER Network offers data transmission with guaranteed quality of service and virtual private networks enabling the usage of advanced applications and services.

Backup links as well as 24h/7d monitoring of the Network Operation Centre ensure high reliability of the network.

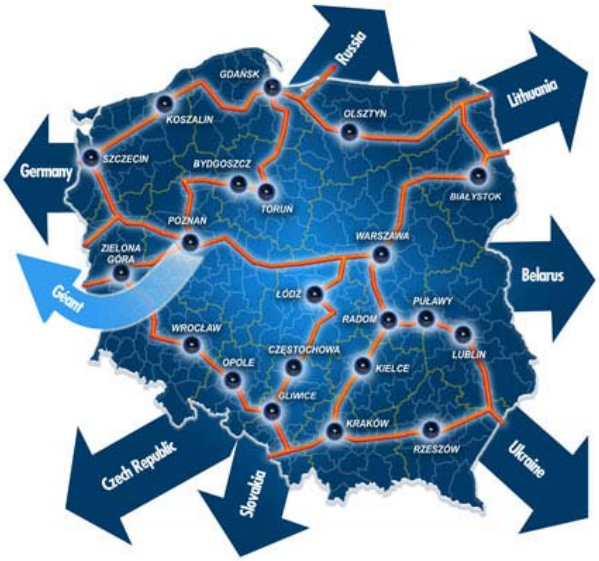


**Figure 11 - PIONIER network**

## 3.4.11. Spain Infrastructure: RedIRIS

RedIRIS is the national research and education network (NREN) for Spain. It is part of Red.es, which also administers and oversees the .es national Top-level domain. Most Spanish universities and research centers are interconnected through RedIRIS, currently totalling about 260 institutions. RedIRIS also acts as an Internet Service Provider for affiliated institutions, through links with Telia and Global Crossing. As a national NREN,

RedIRIS is connected to the high-speed European GÉANT2 backbone, similar to the US-based Abilene Network (see Website: http://www.rediris.es/).

RedIRIS is made up by a set of nodes distributed throughout the country, which are interconnected by a mesh network with a 2.5 Gb core (see Figure 12).

At present there are 18 nodes, one in each Autonomous Region. A node is a set of communications equipments that concentrate the backbone transmission media and access lines of the centres of each region. All these equipments are configured and managed by RedIRIS Network Operation Centre.

Other services provided by RedIRIS include a collection of FTP mirrors, customized software for affiliated institutions, interconnection with the Eduroam European wireless network, LISTSERV service, BlackList IP Service for relays, and an incident response team, IRIS-CERT.

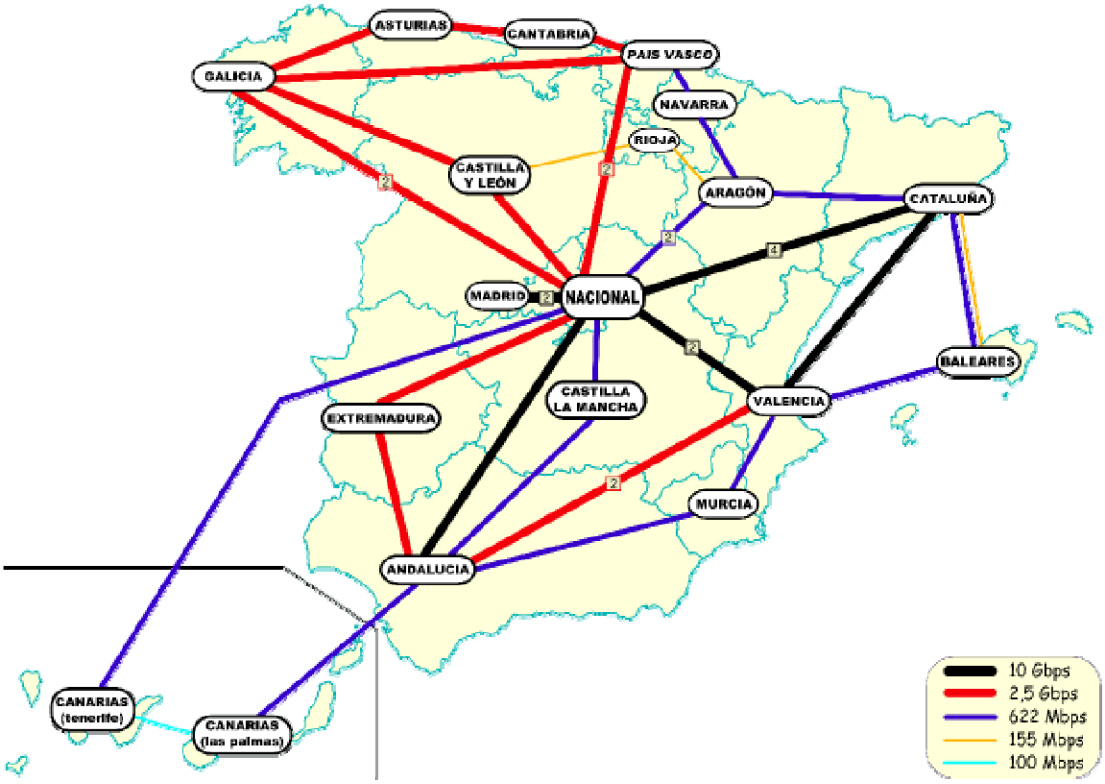


**Figure 12 - RedIRIS infrastructure**

### 3.4.12. French Infrastructure: RENATER

Renater (Réseau national de télécommunications pour la technologie, l'enseignement et la recherche) is a French national public high speed computer network dedicated to connecting universities, schools (including colleges, high schools, engineer and business schools, etc.), and research centers through the whole country. It is funded by the French government (see Website: http://www.renater.fr/).

It connects more than 1000 locations through a high speed, highly meshed network (access up to 2.5 Gbit/s, with a core network up to 80 Gbit/s).

Renater is connected to international networks through a 10 Gbit/s link to GEANT and directly to northern America through two 2.5 Gbit/s links (see Figure 13).

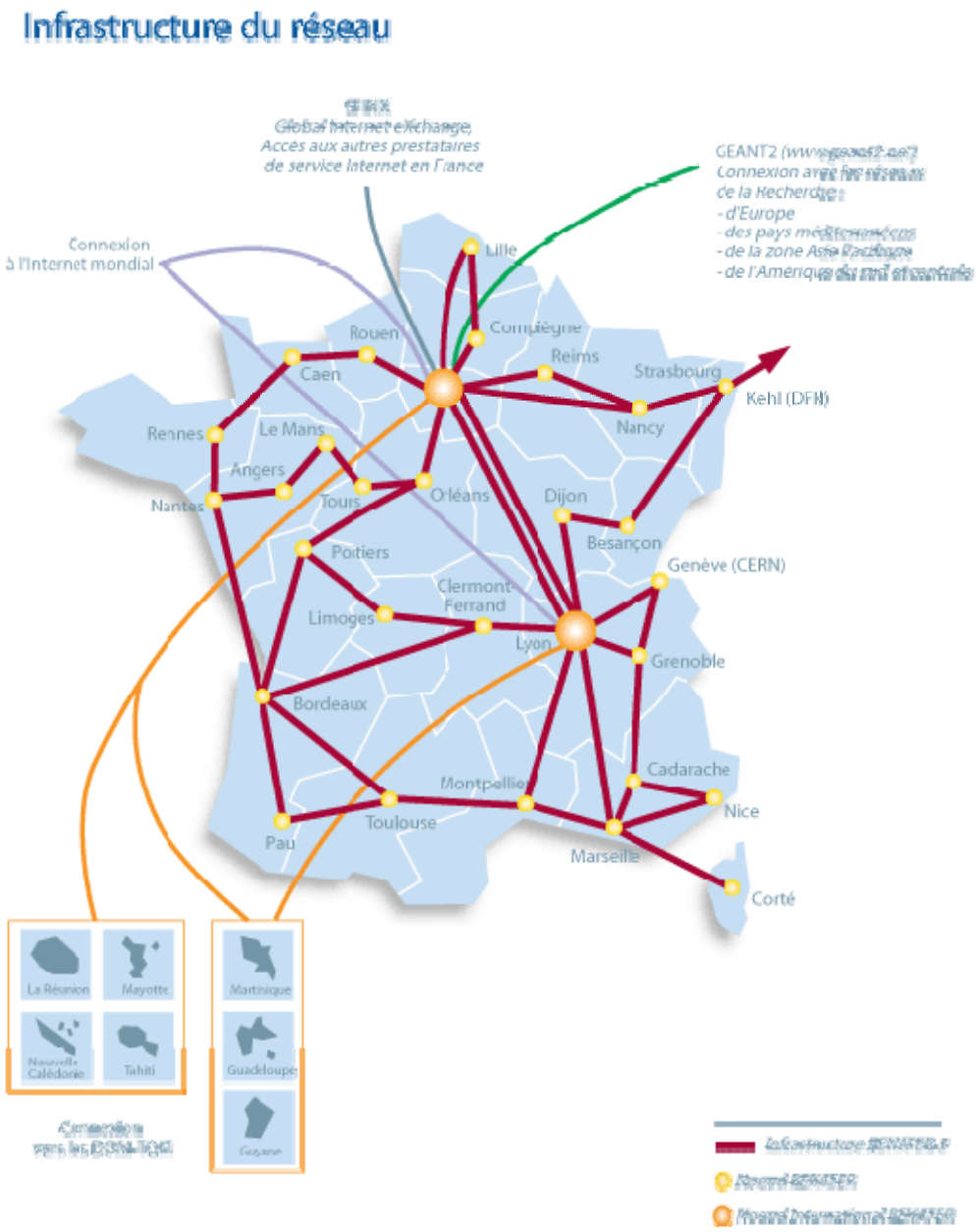


**Figure 13 - RENATER infrastructure**

### 3.4.13. Slovakia infrastructure: SANET

The SANET network is built in frame of SANET2 project and at this time it covers 23 towns (see Website: http://www.sanet.sk). The network infrastructure is based on leased dark fibres, which are terminated in Cisco Catalyst gigabit Ethernet switches.

SANET Network is configured as two rings providing full redundancy with a maximum delay of 5ms. The national connectivity is realized through the Ethernet link in to Slovak Peering

Center SIX placed in Computer Center of the Slovak Technical University (CVT STU) in Bratislava. Its speed is 10Gbps.

The foreign connectivity (see Figure 14) is realized through leased dark fibre at the speed 10Gbps to ACOnet node in Vienna (AT), CESNET node in Brno (CZ), PIONIER node in Bielsko-Biala (PL) and local link to GEAT PoP in Bratislava (1Gbps) and GTS PoP in Bratislava (1,2Gbps).

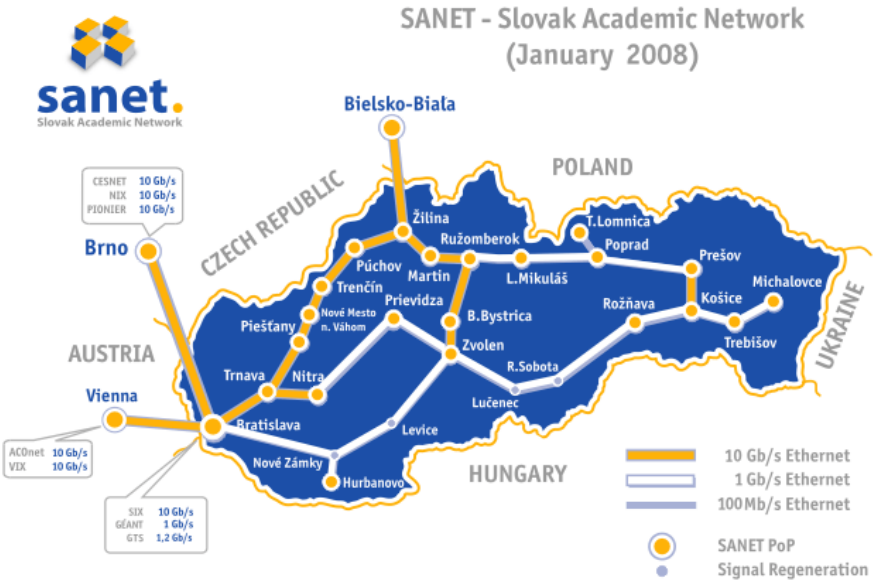


**Figure 14 - SANET network**

# 4. Data Repositories

## 4.1 Introduction

The data repositories which are presented here correspond to fully answered forms received by the contacted people. Each answered form is presented in the document as follows:

- An introduction: presenting a description and the link to the public web site;
- Archive systems: archive backup and recovery systems, archive media in use (transcription systems and policy);
- Archive access and connection: via gateways, or high speed networks;
- Archive content (EO missions, duration, coverage and volume);
- Data curation and long-term preservation (strategy, standard followed, …).

GENESI-DR partners (ASI, CNES, ENEA, DLR, ESA, INFOTERRA, JRC, KSAT and NILU) have also been directly contacted so that they provide the same information about their digital repositories.

## 4.2 ASI Data Repository

### 4.2.1. Introduction

The Italian Space Agency came into being in 1988. ASI has become one of the most significant players in the world in space science, satellite technologies and the development of mobile systems for exploring the Universe. ASI has three operative centers in Matera, Trapani and Malindi, and a Data Center in Frascati, near Rome.

The partners partaking to GENESI-DR come from the Matera center (or Space Geodesy Center). Their archive system is peculiar, as described hereafter.

### 4.2.2. Archive systems

ASI has two sources of data which are:

- ESA data at IPAF/IPAC, holding ERS and ENVISAT (ASAR+MERIS) data;
- ASI data, consisting in XSAR (SRL1/2 + SRTM) data (images and DEMs).

The ESA IPAF (Italian Processing and Archiving Facility)/IPAC (Italian Processing and Archiving Centre) acquires, stores, process and distribute data to users, but do not expose a catalogue to user neither accepts orders from users. In fact, it acts as a provider of data storage and data processing services towards the Central PAF/PAC located at ESRIN. Users ask for registration to such Central PAF/PAC and use the discovery and catalogue services using EOLi (Earth Observation Link), a system which is NOT under the IPAF/IPAC control. It is up to ESA-ESRIN to gather requests from users and send production orders to IPAF/IPAC. Distribution of data to users is done or with media or even electronically, sending authorization info (user, passwd) to the user by phone or email.

For EOLI, see Website: http://earth.esa.int/EOLi/EOLi.html.

As for ASI Archive, it is made up of four resources:

- A Data Archive host (SIR-C/X-SAR SLR 1 and SLR 2 missions);
- A Web server host;
- A Firewall host;

- An Applications Service Provider host (for example, the two following applications integrated within GENESI-DR: QUITSAT, PRIMI).

The archived data is on a Dell Server PE 1900 with four Hard-Disks of 4TB configured in RAID 5. All others hosts are Dell Server PE 1900 with a 1 TB Hard-Disk.

For all hosts, the operating system is Scientific Linux 5.1 hosted into a OpenSuse 11 OS through VirtualBox application.

### 4.2.3. Archive access and connection

Access to the archive is through an HTTP/HTTPS server.

The Internet connection is provided by GARR consortium at 8 mbit/second. Consortium GARR was established by the national Academic and Research Community. Its member institutions are the Italian Universities and the main governmental Research (such as CNR, ENEA, …). It interconnects over 250 universities, laboratories, libraries, observatories and other research facilities. GARR is connected at 10 Gbps to GEANT and to the North American and Japanese networks (see Website: http://www.gar.it).

For security, there is a firewall machine through which the Web archive is accessed.

### 4.2.4. Archive content

The Archive is made of SAR (Synthetic Aperture Radar) image data from the two SIR-C/X-SAR SLR1/SRL2 missions. The Shuttle Imaging Radar-C/X-Band Synthetic Aperture Radar (SIR-C/X-SAR) is an imaging radar system scheduled for launch aboard the space shuttle in the mid 1990s with two planned reflights in different seasons.

The two Data Series are about the flights of the shuttle Endeavour on STS-59 April 9-20, 1994 and on STS-68 September 30-October 11, 1994.

The datasets are archived on-line. Their metadata are accessible through any GIS Client (such as QGIS). Moreover, they can be downloaded from the GIS client itself.

The Archive features are resumed in the Table 2:

| **Archived data sets features** | | | |
|---|---|---|---|
| **Dataseries** | **Number of datasets** | **Dataset File Format** | **Total archive size** |
| SIR-C/X-SAR SLR1 | 571 | ZIP | 11.6 GBytes |
| SIR-C/X-SAR SLR2 | 690 | ZIP | 13.2 GBytes |

| **Detailed features for data series** | |
|---|---|
| **File name** | **File content** |
| DOPPLR.TXT | ASCII file with calibration data |
| DTHDR.TXT | ASCII file with several metadata about time, date, coordinates of acquired data |
| QA.TXT | ASCII file with quality data |
| SEG_xx.IMG | The data itself, that is “Quick-Look” SAR Image data on a PIXMAP not standard format |

**Table 2 - ASI data sets and series**

### 4.2.5. Data curation and long-term preservation

The only data preservation strategy is based on the two following aspects:

- All the original data of the two SIR-C/X-SAR datasets are on a CD-ROM archive, so the on-line archive is a perfect copy of the off-line CD-ROM media archive;
- Data on-line is accessible from everywhere and their metadata can be queried through a Web GIS Client.

Apart from that, until now, there has not been implemented any other kind of data curation and long-term preservation strategy.

## 4.3 BSH Data Repository

### 4.3.1. Introduction

The BSH - Bundesamt für Seeschifffahrt und Hydrographie – is Germany's central maritime authority whose wide range of multidisciplinary tasks includes the protection of the marine environment and furthering of sustainable use of the oceans. Meeting the growing demand for marine and geospatial data is an important part of their tasks. The information in this section is compiled from: http://www.bsh.de/en/index.jsp and the questionnaire returned by Friedrich Nast (BSH).

### 4.3.2. Archive systems

The actual data amount is about 100 GB with an expected growing of 20 GB, with data archived on tapes.

### 4.3.3. Archive access and connection

BSH operates a geo data portal GDI-BSH. Users can access processed Geo- data in a map context via den Map Viewer. This service is also able to link and compare data sets of different information systems of BSH.

There is also the possibility to use OGC services - Web Mapping Service (WMS) and Internet Mapping Service (IMS).

### 4.3.4. Archive content

BSH archives especially data sets acquired by research vessels. It archives maritime and oceanographic data sets from 1873 till today.

### 4.3.5. Data curation and long-term preservation

For long term preservation BSH uses backup strategy and redundant systems. Data sets are stored in accordance with IOC/ICES. Corresponding Metadata fulfil ISO 19115:2003.

## 4.4 CarboEurope-IP Data Repository

### 4.4.1. Introduction

CarboEurope-IP aims to understand and quantify the present terrestrial carbon balance of Europe and the associated uncertainty at local, regional and continental scale. This means to:

- determine the European carbon balance with its spatial and temporal patterns;

- understand the controlling processes and mechanisms of carbon cycling in European ecosystems and how these are affected by climate change and variability and human management;
- develop an observation system to detect changes in atmospheric CO2 concentrations and ecosystem carbon stocks related to the European commitments under the Kyoto Protocol.

To advance our understanding in a multidisciplinary and integrated way, 61 research centres from 17 European countries have joined forces for a 5-year EU-funded research project started in January 2004. See repartition here below in Figure 15:

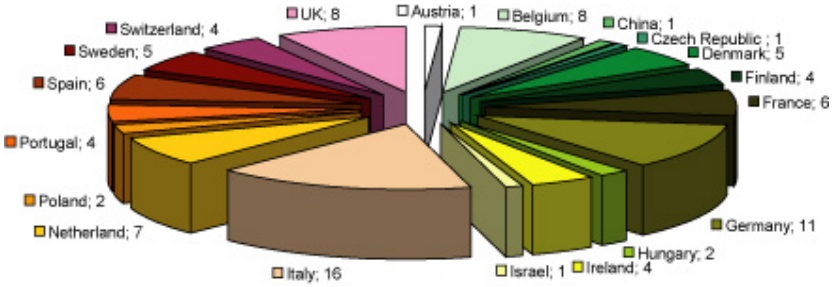


**Figure 15 - CarboEuropeIP Institutes repartition**

CarboEurope-IP addresses basic scientific questions of high political relevance.

Many PIs (Principle Investigators) contribute different data. Below, the ecosystems repository of the DISAFRI University of Tuscia, Italy is specified.

The information in this section is collected from the questionnaire returned by Mr Pete Smith and from the Website http://www.carboeurope.org/.

### 4.4.2. Archive systems

The University system is used (IBM server with SQL). It is a non-redundant back-up system, and an off-line copy is stored at HD.

### 4.4.3. Archive access and connection

External users are granted access via web-site portal. Registration is needed, and the data are available after authorization by the owner depending on site and year (part of the data are public after registration). The repository has access via a gateway and GARR- network. An external process capacity is provided, for quality control and gap filling of Eddy data.

### 4.4.4. Archive content

For instance, data from Eddy covariance towers (10GB), and expected 1-2 GB per year.
Data standards are according to the internal policy.

### 4.4.5. Data curation and long-term preservation

Volatile information within the meta-data is updated by the PIs and a record is kept. The versions are tracked by standard processing.

## 4.5 CLS Data Repository

### 4.5.1. Introduction

CLS, a subsidiary of CNES (French Space Agency), IFREMER (French Research Institute for exploration of the sea) and several French financial institutions, offers satellite services in location, environmental data collection and ocean observations to a broad range of professionals including: government, industry and the scientific community.

Within this context, CLS works in close collaboration with CNES, the National Oceanic and Atmospheric Administration (NOAA) of the United States, Eumetsat (Organization for the Exploitation of meteorological satellites), JAXA (Japan Aerospace eXploration Agency) and INPE (Instituto Nacional de Pesquisas Espaciais).

The information in this section is collected from the questionnaire returned by Mr Blanc and from the CLS web site (http://www.cls.fr).

### 4.5.2. Archive systems

The archive system is on an ATOLL platform and is using the CNES STAF service (dedicated to the archiving, storage and retrieval of data).

### 4.5.3. Archive access and connection

The archive is accessible through a web site portal depending on the project:

- http://www.mersea.eu.org for Mersea (Marine Core Service for GMES);
- http://www.aviso.oceanobs.com for SALP/AVISO (Satellite altimetry);
- http://www.mercator-ocean.fr for Mercator Ocean (Ocean forecasting center);
- http://www.seadatanet.org for Seadatanet (Sea Data Network).

The access is free for historical data and restricted for current data.

### 4.5.4. Archive content

The data managed by CLS comes from several sources: satellite, in situ, ocean modelling, statistics, …

The amount of data depends on the project.

The covered period starts more than 20 years ago for the older project to a few months for the last ones.

### 4.5.5. Data curation and long-term preservation

The strategy for data curation and long-term preservation is the same as the CNES one (use of STAF service).

## 4.6 CNES Data Repositories

### 4.6.1. Introduction

The CNES – Centre National d'Etudes Spatiales – is the French space agency. There are no general rules to manage each Earth Observation Data mission: it depends on the specific collaboration.

However in the following sections the major strategic axes of CNES Toulouse for Earth Science.

### 4.6.2. Archive systems

As for the data stored at CNES a common archiving system is available: the STAF (File Transfer and Storing Service) upon which missions can rely for the storage and management of their data products.

The following Figure 16 shows how STAF is used:

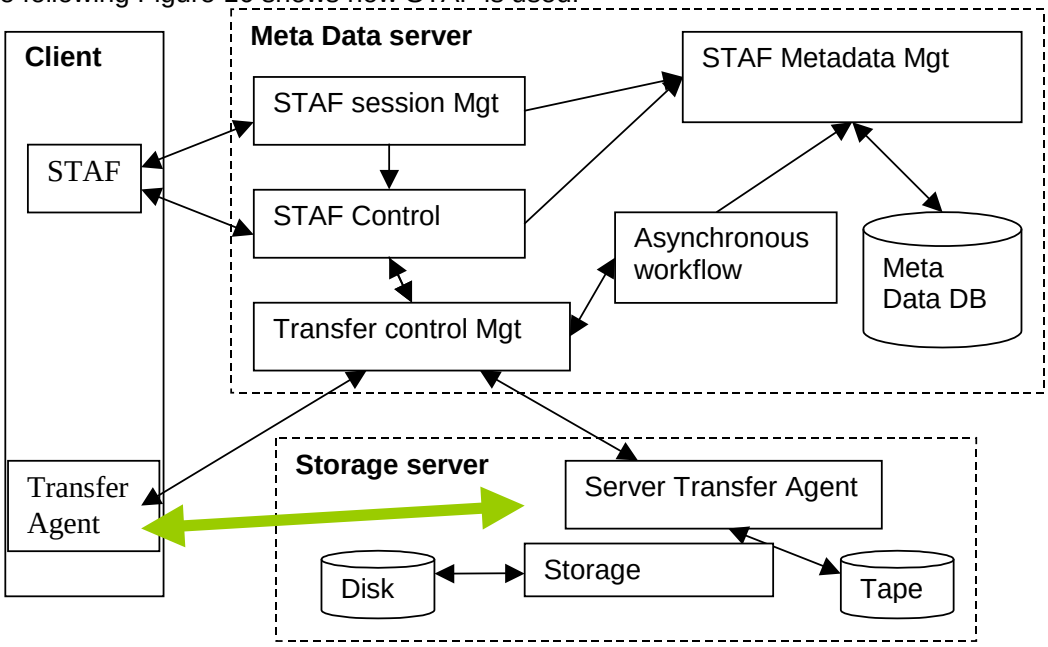


**Figure 16 - CNES STAF organization**

### 4.6.3. Archive access and connection

CNES uses SIPAD (Système d'Information, de Préservation et d'Accès aux Données or Information System for Data Preservation and Access) for managing most of its mission data. Data centers such as for example, ICARE, SALTO, MERCATOR also use SIPAD.

Indeed, SIPAD is a generic software system allowing web consultation of scientific data catalogs and data access. It is adaptable to any scientific domain and, therefore, can be used by Data Centres to implement their data access function without having to develop it from scratch.

It offers the following functionalities (see Figure 17):

- Mechanisms for searching for relevant data;
- Data selection and ordering;
- User and access right management.

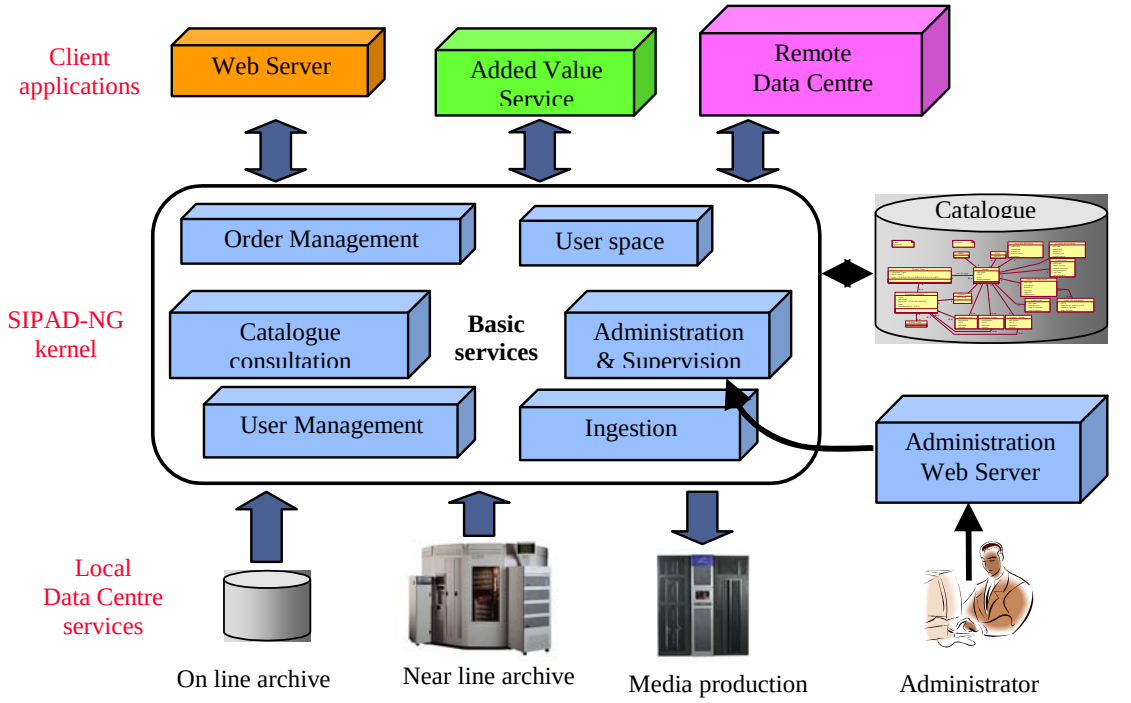


**Figure 17 - CNES SIPAD**

### 4.6.4. Archive content

The archive contains data form different projects: SPOT, Mercator, soho, POLDER, SALTO… . The STAF system is used to archive those data.

### 4.6.5. Data curation and long-term preservation

A new project called SERAD (Services de Référencement et d'Archivage de Données or Data Recording and Archiving Service) is involved in long-term preservation of archived data at CNES. Its objective is to settle a central data system for all the space data under CNES responsibility. It uses CNES means: STAF, SIPAD for data access and BDMS (Bureau des Metadata et Services or Metadata and Services Register) for metadata access. It also used the iDN (International Directory Network) which is an international server of data set descriptions, maintained by NASA in the framework of CEOS for space agencies.

## 4.7 CNIG Data Repository

### 4.7.1. Introduction

The objectives of the CNIG (Centro Nacional de Informacion Geografica) are :

- To commercialize and to distribute the geographical products elaborated by the National Geographical Institute, and for other organs and organisms of the General Management of the State.
- To divulge and to inform about the availability and characteristics of the existing geographic products in order to be used by public organisms, enterprises and private individual.

- To provide advice and technical backup to other organism and centres of the Public Administration, in subjects like cartography, geodesy, geophysics sing, and the introduction and development of Geographic Information System (GIS).
- To establish commercial agreements with editorial sector and computer enterprises, in order to develop specific products using National Geographic Institute's own.
- To increase the value of basic data produced by National Geographic Institute, adapting them to the clients requests.
- To strengthen the presence of National Geographic Institute - National Centre of Cartographic Information into international organisations.

The information in this section is collected from the questionnaire returned by Pedro Vivas White and from the CNIG web site.

### 4.7.2. Archive systems

The data is stored on SDLT-IV equipment.

### 4.7.3. Archive access and connection

The access is made through the web site portal (http://www.cnig.es/) with free and registered users access.

### 4.7.4. Archive content

The archive contents data from aerial and satellite sources for mapping activities. The amount of data is about 50 GB. It is updated once or twice a year.

### 4.7.5. Data curation and long-term preservation

Data curation and long-term preservation is considered by an internal policy, based on backup, redundant and recovery systems.

## 4.8 CoreH2O Data Repository

### 4.8.1. Introduction

The CoreH2O mission is one of the six Core candidate missions selected by ESA and aimed at covering a range of environmental issues with the aim of furthering our understanding of the Earth system and changing climate. CoreH2O mission aims to provide accurate information on the characteristics of the cycle of snow, glaciers and surface water. It includes the use of a Synthetic Aperture Radar (SAR) with two frequencies (9.6 and 17.2 GHz) to be employed in two consecutive phases to obtain all-weather, year-round information on regional and continental scale, on snow and water (see http://www.esa.int/esaLP/SEM1J1C1S6F_index_1.html).

The Remote Sensing group of the Institute of Meteorology and Geophysics at the University of Innsbruck (http://imgi.uibk.ac.at/remotesensing/intro) is involved with the CoreH2O project with their project : Ice/Climate interactions on Larsen Ice Shelf, which studies glacier response to climate change.

### 4.8.2. Archive systems

Currently the Remote Sensing group of the Institute of Meteorology and Geophysics is using the central archiving facility of the University of Innsbruck, which is a redundant system and recovery system with backup copy, media_store_copy on DVD.

### 4.8.3. Archive access and connection

The University of Innsbruck archiving facility is used for archive access and connection.

### 4.8.4. Archive content

Data are provided from ESA and DLR and stored and used for project work.

The sensor proposed for CoreH2O is a dual frequency SAR, operating at 9.6 and 17 GHz, VV and VH polarizations. Preparatory activities for the mission include the development and improvement of theoretical models for radar scattering of snow, and the collection of backscatter signatures under various snow conditions in test sites in Europe and North America. Various backscatter models have been developed that are applied for signature studies and development of algorithms for snow parameter retrievals. For volume scatter modelling the selected methods include dense medium radiative transfer and the wave approach with scattering characteristics being calculated for small unit volumes applying the discrete dipole approximation.

The dual frequency, dual polarization approach of CoreH2O enables the decomposition of the scattering signal into surface and volume contributions and the estimation of snow optical thickness and scattering albedo. This is one of the baseline versions for retrieval of snow water equivalent, a key parameter for snow hydrology and climate research. The method is applicable for dry snow packs that are characterized by comparatively low dielectric losses. Ongoing work is concerned with validation of the theoretical models by means of experimental data, studies of backscatter sensitivity to snow physical properties, and development of retrieval algorithms.

### 4.8.5. Data curation and long-term preservation

The University of Innsbruck archiving system is relied on.

## 4.9 DLR Data Repositories

### 4.9.1. Introduction

The German Remote Sensing Data Centre (DFD) is an institute of the German Aerospace Centre (DLR). Together with the Remote Sensing Technology Institute (IMF), it comprises DLR's Applied Remote Sensing Cluster, in which are concentrated many of DLR's activities related to earth observation with satellites and aircraft (see Figure 18).

DFD is concerned with the reception, **archiving**, **distribution** and utilization of data. In addition to applied research objectives, DFD is expert in the development and operational performance of information technology systems. Beside a multitude of other areas the technical expertises is also concentrated in Long-term data archiving in the National Remote Sensing Data Library. Via the Internet this data can be accessed, ordered, and in some cases directly downloaded.

As a DLR research institution, DFD is firmly embedded in European and international space infrastructures. It acts on behalf of the European Space Agency, ESA, as a data and processing centre for European and international earth observation missions. It is also a partner of NASA in other missions.

DFD is currently establishing a World Data Centre for Remote Sensing of the Atmosphere on behalf of the International Council of Scientific Unions.

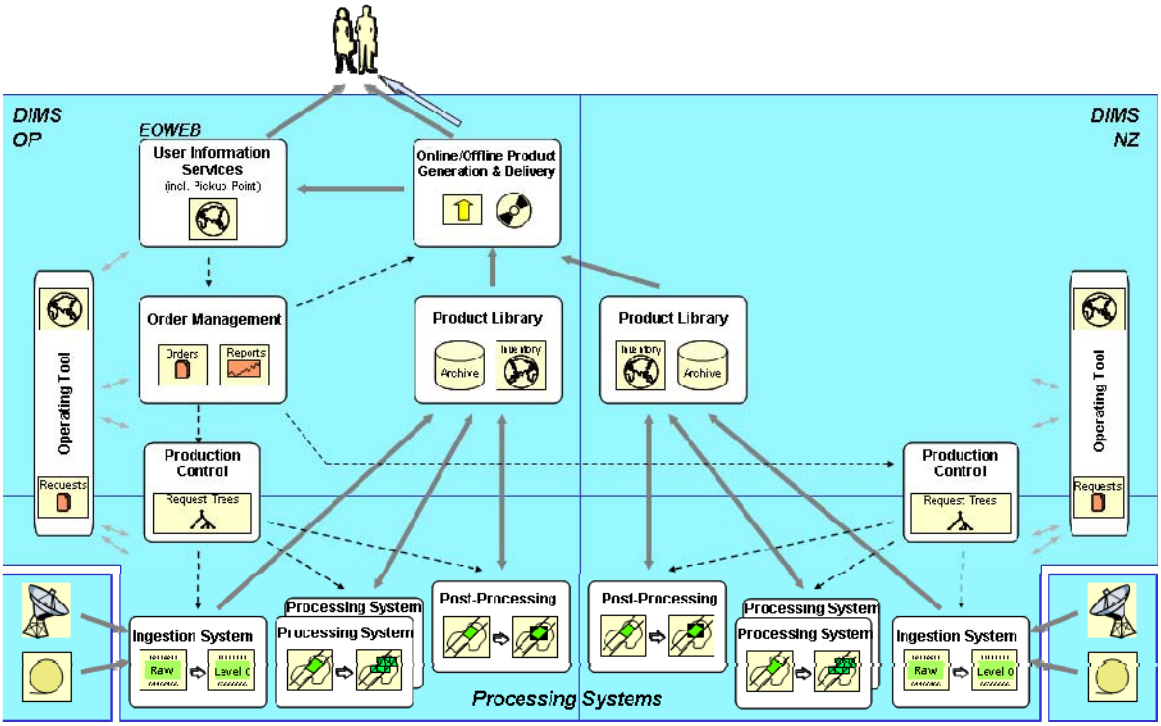


**Figure 18 - DLR's Data Information and Management System (DIMS)**

### 4.9.2. Archive systems

At present DLR operates its distributed hierarchical archive system (in Oberpfaffenhofen and Neustrelitz, as shown in Figure 18) on basis of several technologies, like Quantum AML/2, Quantum i2000, Sun Solaris, Sun SAM-FS, IBM Informix, Progress ObjectStore. Currently storage media are STK9940 and LTO-x.

### 4.9.3. Archive access and connection

External users can access data via web site portals DLR EOWEB and DLR EOWEBÂ® - Next Generation. A running future project is the provision of data via interoperable service protocols such as CEOS/CIP and ESA HMA/OGC. The applicable access policy (also for raw data) is depending from mission. In general there is a distinction between free/registered users, and selected users with password.

The adherence of access policy is realized by several firewalls and access control. The archives are connected to the external world by the “Deutsches Forschungsnetz” – Germany’s National Research and Education Network (600 Mbit/s), HISEEN/GEANT and EUMETCast/GEONATCast.

### 4.9.4. Archive content

The archive system contains 0.5 PB of satellite data from more the 20 sources organised in 12 million files (above all satellite images incl. auxiliary data). Archive history goes back to 1980ies. The yearly growth is expected by 0.5 PB.

### 4.9.5. Data curation and long-term preservation

Data are described by a metadata system. Each data set exists in at least 2 copies in redundant systems. Data sets are migrated regularly.

## 4.10 DLS Data Repository

### 4.10.1. Introduction

The Department of Lands and Surveys (DLS) is a department of the Republic of Cyprus. It manages the major land matters of Cyprus regarding surveying including maintenance of the state survey infrastructure, mapping, investigation into title, registration, conveyance, valuations, conservation and management of State Lands, compulsory acquisition/requisitions and encumbrances. Beside other tasks the Department acts as the Official Cartographic Organization of the Republic for the compilation, production and issue of maps and plans.

The information in this section is compiled from:

http://www.moi.gov.cy/moi/dls/dls.nsf/dmlindex_en/dmlindex_en?OpenDocument

and the questionnaire returned by Andreas Hadjiraftis (DLS).

### 4.10.2. Archive systems

The current archive system is based on following technologies: Unix Sun Solaris, Windows XP/Vista, Oracle, ArcGIS8.2 and ArcGIS 9.2, Arc/Storm, and ArcSDE. Data archived on tape (data cartridges).

Basically, all Legal/Fiscal application systems using the case tools of ORACLE Relational Database Management System. The spatial applications are based on Arc/Info, Arc/View and other ESRI GIS products and RDBMS technology, on surveying packages (such as LISCAD) and on CAD packages (eg. AutoCAD and MicroStation).

A number of application systems have been developed around the system. These systems basically include applications for data entry, maintenance, storage, enquiry and output (displaying, reporting, plotting and printing).

### 4.10.3. Archive access and connection

An access to external users via gateway is possible. Appropriated tools are necessary. Data are password protected. The department also provides raw data on physical media.

### 4.10.4. Archive content

The Department records a considerable amount of land related data. Information about development, utilities, land use, water resources, geology, and even statistical data for population, industry, agriculture and planning, are recorded on, or closely related to the cadastral plans.

Four main databases are operated in the Department:

- The Survey Database;
- The Digital Cadastral Database;
- The Topographical Database;
- The Legal/Fiscal Database.

The databases contain the survey data-set, the digital cadastral data-set, topographical overlays, the legal/fiscal data-set, sales history data, and other useful information.

### 4.10.5. Data curation and long-term preservation

Data curation and long-term preservation is considered by an internal policy, based on backup, redundant systems and recovery systems. DLS performs regularly upgrading of hardware and software and staff training.

## 4.11 ENEA Data Repositories

### 4.11.1. Archive systems

Two different systems have to be described as ENEA DR: the ENEA MOON VOS and the CNR GOS SST.

ENEA is actively involved in the operational monitoring of the Mediterranean Sea with ships of opportunity. This is part of an international IOC – WMO program as well as part of the Mediterranean Forecasting System. For this application, there is the obligation to provide access to data in near real time. The entire data acquisition – management system is called MOON VOS.

There are two kind of users for the MOON VOS system:

- the operational users, who must access the data for their assimilation in forecast models;
- researchers and public who are using data for climatic or environmental studies.

The operational archive is composed by a directory of data files arranged at a daily base, that are included in a server http accessible.

The architecture developed for the MOON-VOS portal is based on the idea that the software must be freely available in the net and the re-use must be assured in many fields of application. It includes the following elements: Bus, Right Management, Registry, Network Services.

The bus is the software allowing the connection of portals and applications, by means of ad hoc developed interfaces. In our case, two data bases are accessed: a local one containing *in situ* data and satellite data residing in a server located remotely. Both data were collected in the framework of the Mediterranean Forecasting System.

The Right Management allows the management of authentication, authorization and logging.

Registry is based on a governance model (mission, quality and business) that allow the data maintenance, correct processing and interpretation.

The network services are the set of services provided to users, such as discovery, view, download, transformation. In particular, discovery allows the search of data through the metadata in catalogues, view allows to view at maps of data layers, download allows to access the data selected through the discovery service.

Information is included in a relational data base, with tables that can be schematized with the following scheme:

- Metadata tables: Parameters (Gf3), Roscop type, Matrix and environ. Compartment and Quality flag decode;
- Administration tables: User and permission, Log table, Utility tables and Cache table;
- Data Tables:
    - Table for X,Y,Z,T station limit;

- Table for Cruise metadata (data protection included);
- Pivot table to control : Table X,Y,T constant (Xbt i.e.), Table X,Y,Z constant (current meter i.e.), Table Z constant (drift i.e.), Table generic (X,Y,Z.T variable), Table for Image, Station parameter, Quality flag;
- Table of header (to trace original header data);
- Table of note (comments).

This scheme allows the organization of information in a flexible way, extending the number of parameters in each station and optimizing the storage. This is done by the use of pivotal tables containing data having some similar characteristics. For example, the vertical profiles are constant in longitude, latitude and time, the table related to a such stations contains only the measured values of each parameter and their quality flags. The same procedure is applied for temporal series or other kind of stations. This procedure allows a faster response of the information system to queries.

Other tables or registries include: the list of parameters and measurements type, all units of measurement used, all code lists/thesauri used in the application schema, identifier namespaces, all feature catalogues, metadata on dataset level will be available through catalogue services and registry for the management of right access.

Data access is organized in three levels (public, limited, confidential) on the bases of wishes of the data owner. All ‘stations’ belonging to a defined ‘cruise’ have the same level of confidentiality.

The CNR GOS SST archive is composed by daily images at full resolution (1km) for the entire Mediterranean. The data are organized in files and included in a THREDDS server.

### 4.11.2. Archive access and connection

The operational daily files are accessible through http protected by password

The archived files for other users are in a MySQL data base and data can be accessed via internet services.

The development framework is done to achieve user-friendliness by allowing users to express intuitively their preferences in ordering products. The logical hierarchy of the information system is taking into consideration that the users is starting his queries by selecting geographical and temporal windows. He wants to look at the products available in that area and their temporal distribution, to know what parameters are included in the products, possibly have a look at them and finally download the products. This logical hierarchy has been used to develop the interfaces that allow the discovery, selection and view of products.

The functional scheme is shown in the following Figure 19:

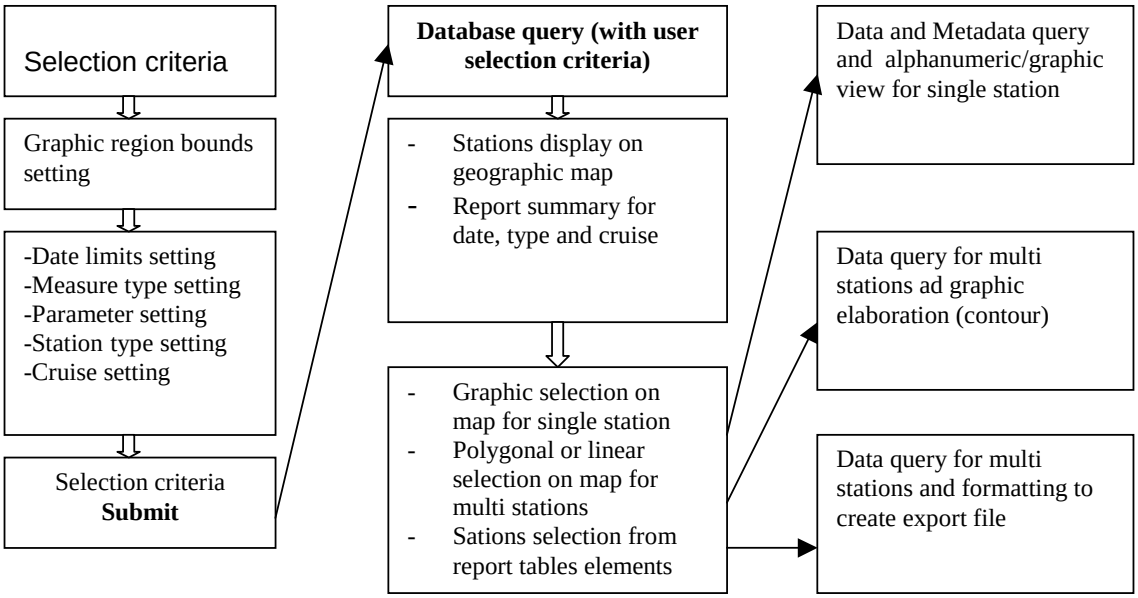


**Figure 19 - ENEA functional scheme for discovery, selection and view of products**

The functions are viewed by the users through an Internet browser. The graphical interface is divided in frames allowing (see Figure 20):

- the selection of the stations on the base of cruises, parameters, geographical location, time;
- the response frame providing the view of the selected stations; the selection can be refined in this frame;
- display of selected stations by query;
- other links and title frames.

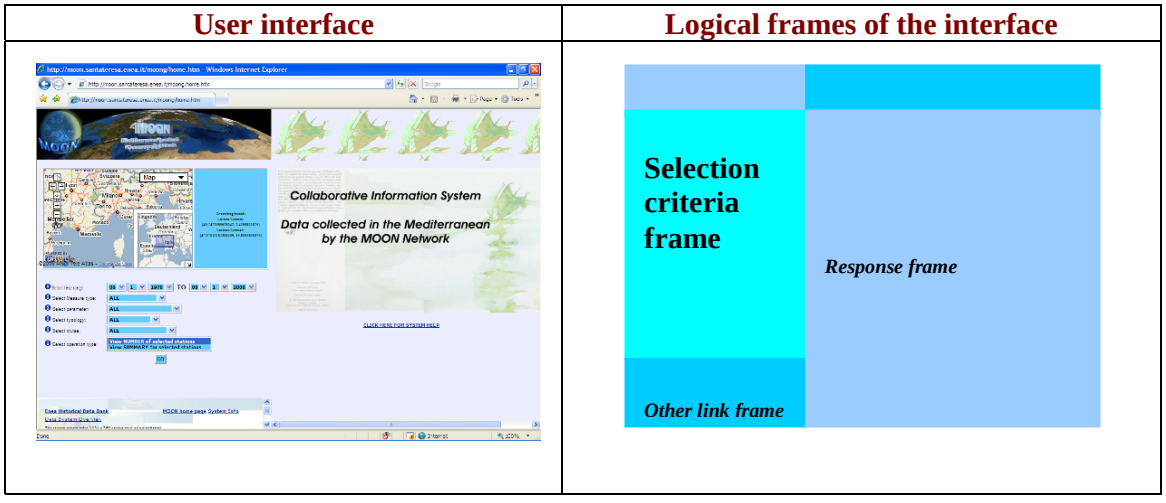


**Figure 20 - ENEA query GUI**

The geographic region can be selected through Google Maps API (Google Maps embedded in web pages with JAVA script; the API provides a number of utilities for manipulating maps). The selected area appears in the 'response frame' with the number of stations contained in that area.

The response to the queries are shown in a Google map together with a table with a temporal distribution of selected stations. In the 'query stations data frame' appears a table from which it is possible to look at the metadata, profiles or data numbers.

Other view services developed for the users are the horizontal and vertical maps that can be done 'on the fly' respectively with polygons or polylines.

Finally data can be exported in two formats: ASCII MedAtlas and ODV (Ocean Data View) spreadsheet. The ENEA information system is also starting to produce data in netCDF format.

View and download services are provided under the data policy decided by products' owners.

The CNR GOS SST data are accessible through a THREDDS catalog.

### 4.11.3. Archive content

The archives contain in-situ temperature profiles and satellite SST (Sea Surface Temperatures).

### 4.11.4. Data curation and long-term preservation

The data are maintained in the digital repository. Their quality is controlled from time to time, in order to assure the consistency of the different data sets.

Data are also used for re-analysis of the operational systems, as is usually done by the meteorological services.

Finally the data are used also for the investigation of the climate and environmental variability.

The curation is involving a periodic quality control, the validation of the different data bases, the long term archiving for scientific studies.

## 4.12 ESA Data Repositories

### 4.12.1. Introduction

ESRIN, known as the ESA Centre for Earth Observation, is the ESA establishment responsible for managing the operation and exploitation of ESA's Earth Observation satellites. In cooperation with other space agencies, it also manages the acquisition, distribution and exploitation of data from non-ESA satellites. The world's largest database of environmental data for both Europe and Africa is managed from ESRIN.

In carrying out this work, ESA's Earth Observation Directorate works closely with national space agencies, both in ESA Member States and worldwide, as well as with coordination and standardisation bodies. It also cooperates with many small-and medium-sized enterprises, and with the service industry.

Data from the many instruments on board ESA satellites, and from more than 20 non-ESA Earth Observation satellites, are sent to a network of worldwide acquisition stations, processed in these stations or at archiving centres, and then distributed to a worldwide user community that includes several thousand scientists, value companies or application centres.

Instrument performance and product quality are permanently checked and new products developed in response to evolving user demand. Responsibility for ensuring this is done quickly and efficiently lies with those working in Earth Observation at ESRIN.

### 4.12.2. Archive systems

In 2003 the European Space Agency launched, in line with ESA's Oxygen initiative, a strategy for the evolution of the several Earth Observation (EO) missions ground segments (handled and/or to be developed) into an open multi-mission architecture, which includes as main goals:

- Adoption of a common architecture for all missions;
- Decomposition of the facility architecture into functional block elements;
- Harmonization and standardization of interfaces;
- Evolution of current missions payload data segments into the common architecture;
- Re-utilization of already available and tested elements;
- Development of a generic multi-mission infrastructure where the elements specific to each EO mission can be plugged in.

For this purpose, the Agency has already or is in the process of harmonizing within one coherent frame some of the facilities basic functional elements as long-term and on line archive, order handling, systematic processing, product distribution including online delivery and inter-facility product exchange. Support features like monitoring and control and software management complement the framework. The long-term goal of this effort is to define a harmonized European infrastructure.

The resulting architecture, based on the ISO 14721:2003 OAIS (Open Archival Information System) standard (see references [3] and [4]), is named Multi-Mission Facility Infrastructure (MMFI) and forms the common infrastructure over which ESA intends to build its future payload data ground segments.

The MMFI, via its Multi-Mission Facility infrastructure Elements (MMFE), provides data producer oriented services for data archiving, data processor oriented services for data retrieval and processing management and consumer oriented services in support of the ESA central infrastructure services (see Figure 21).

Due to the distributed nature of the ESA's ground segments, composed of many acquisition and archiving centres, a Payload Data Ground Segments for a generic mission can be considered as is composed of:

- a Multi-Mission Central Infrastructure component, consisting of all elements required to provide User Services (cataloguing, user access, data ordering, etc.), and Quality Assurance services (payload data quality control, sensor performance assessment, etc.)
- a distributed Multi-Mission Facility Ground Segment (FGS) component, consisting of all elements necessary for the acquisition, ingestion, long-term archive, order processing and data disseminations to end users of a specific mission. As highlighted before, a generic FGS generically requires elements that are mission-specific like processors and quality control systems, but most of its services can be provided by means of common multi-mission elements. Multi-mission centers typically have one "logical" FGS for each mission operated, but the sub-systems used can be shared if they have built-in multi-mission characteristics.

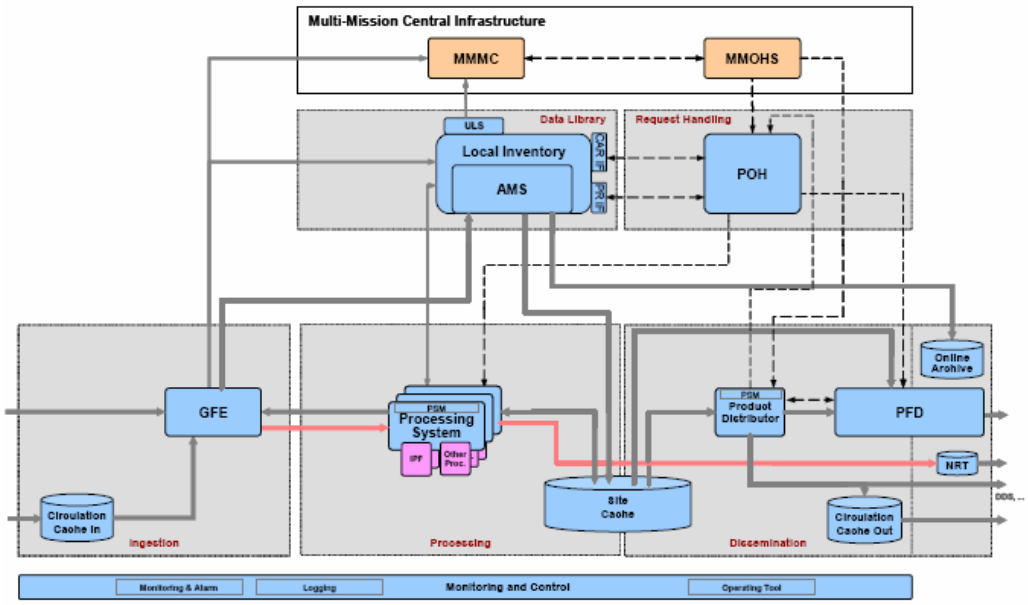


**Figure 21 - Generic MMFI architecture at an ESA center**

The main building blocks of the MMFI are:

- Ingestion - The ingestion workflows are controlled by the "Generic Front End" (GFE) that provides a configurable workflow engine and a standard set of re-usable plug-ins relevant to ingestion. For data product dependent functionality for metadata extraction and browse generation the ingestion assembly relies on the functionality provided by the Data Request Server (DRS) that allows accessing data (including binary data) in a configurable manner.
- Data Library - The data and metadata holdings of a centre are stored in a data library assembly that consists of the data storage component "Archive Management System" (AMS) and an inventory component "Local Inventory" (LI). The AMS manages the actual archiving of the data product and as an abstraction layer allows to change the underlying storage technology without affecting the architecture of the overall system. It is worth noting that, in line with the goal of achieving the maximum harmonization and standardization of the FGSs, a unique storage technology has been adopted by ESA for its EO missions. The migration of ESA's holdings to the new archives is presently on-going.
- Request Handling - Interfacing ESA's Multi-Mission Order Handling System (MMOHS) is a request handling subsystem, the POH (Product Ordering System) that handles production and dissemination requests and organizes the required workflow based on the product type and output medium requested in the order. The POH is supported by a set of auxiliary components that interface other MMFI elements and provide specific functionality for workflow management.
- Dissemination - ESA centres need to support a variety of dissemination scenarios and dissemination channels. The key component for dissemination is the "Product Formatting and Delivery" (PFD), a dissemination workflow management component that can handle different dissemination channels in a configurable manner, and "Product Distributor", a component that manages systematic delivery procedures for data circulation between centers, subscriptions and standing orders.
- Processing - The MMFI support processing by providing a generic "Processing System Management" (PSM) framework. The framework allows integrating mission

and sensor specific processing facilities with minimal effort and offering a choice of protocols for the definition of the MMFI – processing facility interface.

Additional there are utility components for service monitoring, logging and automated monitoring and alarm.

The MMFI is used to implement Facility Ground Segments for all current ESA missions including ENVISAT, ERS and past and current Third Party Missions. The MMFI based ground segments are presently deployed at the ESRIN centre in Frascati (Italy), Matera (Italy) ,Oberpfaffenhofen (Germany), Maspalomas (Spain), Farnborough (UK) and Kiruna (Sweden).

### 4.12.3. Archive access and connection

If data are intended to be used for scientific research, applications development, or research and development in preparation for operational use in future, then user can submit what is known as a 'Category-1' proposal to ESA (see reference [5]). Accredited Category 1 users – known as Principal Investigators – will receive Earth Observation products at cost price. The data are delivered over the internet or via CD or DVD-Rom. To submit a proposal to ESA the Principal Investigator must use the Earth Observation Principal Investigator Portal [http://eopi.esa.int/esa/esa] to supply detailed information about the team involved in their project, the planned use of the data and its scientific background, the requested amount and type of data, and the project schedule. A group of scientists external to ESA will review the scientific merit of the proposal while ESA performs an assessment of data provision feasibility. The Principal Investigator must sign the Terms and Conditions and confirm availability of funding before ordering data. They are expected to report on project progress at least twice a year through the Earth Observation Principal Investigator Portal, and may also publish any intermediate results on the Portal.

ESA also issues regular Announcement of Opportunity for researchers to make use of Earth Observation data within particular fields. Current Announcements of Opportunity can be accessed via the Earth Observation Principal Investigator Portal.

If ESA spacecraft data are required for operational or commercial purposes this is classed as Category-2 use. The Agency has licensed two competing companies - known as Distributing Entities – to resell ESA Earth Observation data and services at market rates. The Distributing Entities are called EMMA and SARCOM, and are represented by Eurimage and Spot Image.

The data, products and services provided by ESA are made available to users via the following key delivery methods (see references [6] and [7]):

- EOLI-SA. The EOLI-SA (Stand Alone) client combines the advantages of both the online and offline catalogues. This tool provides access to the online ESA catalogues of Earth Observation (EO) products. It allows the visualisation of quick-looks and the online ordering of EO products. A copy of the inventory can be downloaded locally to increase response times and to perform interferometric searches on ENVISAT and ERS SAR products. The tool gives access to all collections for which ESA provides products. EOLI-SA is a java application supported on all major platforms (Windows 95/98/ME/2000/NT/XP, Linux, Solaris, Mac OS X and other Unix systems).
- EOLI Web client. The new EOLI Web Client provides access to online ESA catalogues of EO products via a standard web browser, as well as to the catalogues of other data providers (for example DLR and NASA ECHO) and is part of ESA's eoPortal. This tool does not allow online ordering (for which we recommend to use EOLI-SA), but it allows you to browse the meta data and obtain further information about the collections and products available from ESA and other providers.

- DESCW. DESCW is an offline multi-mission catalogue covering ERS, ENVISAT and ESA Third Party missions. It is a tool providing users with the information about satellite/sensor data availability for a given time interval and over a specified geographical location. Orders for products can be formulated using DESCW. Regular updates of the data for use within DESCW are available via FTP.
- IONIA. The IONIA server allows access to the Level 3 Earth Observation Products generated at ESA/ESRIN from ERS and other sources. Ionia is the name of the Greek school of philosophy (VI-Vth century BC) that formulated the theory of the cosmos as constituted by four basic elements (fire, water, air, earth), the same elements that are now observed from space-born instruments and constitute the main fields of interest in the analysis of remote sensing images.
- ATSR Global Fire Atlas. The ATSR Global Fire Atlas provides monthly global fire maps made from ATSR-2 night time data. The maps are generated monthly, and cover the period from September 1995.
- ENVISAT Data Dissemination System - The ENVISAT Data Dissemination System (DDS) is currently providing for rapid dissemination of the satellite's data products to users across Europe. The system uses a commercial satellite-based network, which is based on the Digital Video Broadcasting (DVB) standard and is integrated into the Internet. It accommodates low-cost user stations and allows for cost-effective use of satellite bandwidth. It is also suitable for use in support of other missions.

### 4.12.4. Archive content

ESA provides an extensive range of high-quality satellite data, products and services. Data from several satellite ESA and 3rd Party missions are archived. These include ERS-1, ERS-2, Envisat, Landsat, Proba, QuickSCAT, Terra/Aqua, DMC, SPOT, ALOS, IKONOS, OrbView-2, ODIN, Kompsat-1&2, IRS-P6 [7].

The overall amount of archived data is approximately 4-5Petabytes.

Data are archived in their native format (defined by mission owner). It is planned to adopt Standard Archive Format for Europe (SAFE) format in future (see reference [9]).

### 4.12.5. Data curation and long-term preservation

In 2006, the European Space Agency (ESA) initiated a coordination action to share among all the European (and Canadian) stakeholders a common approach to the long term preservation of Earth Observation space data (see reference [11]). During 2007, the Agency started consultations with its Member States presenting an EO Long Term Data Preservation strategy [9] targeting the preservation of all European (including Canada) EO space data for an unlimited time-span ensuring and facilitating their accessibility and usability through the implementation of a cooperative and harmonized collective approach among the EO space data owners.

The Long Term Data Preservation Working Group with representatives from ASI, CNES, CSA, DLR and ESA was formed at the end of 2007 within the Ground Segment Coordination Body with the goal to define and promote, with the involvement of all the European EO space data and archive owners, the European LTDP Common Guidelines (see reference [10] and [11]) and also to increase awareness on LTDP. The resulting draft LTDP guidelines were reviewed by all ESA member states in the DOSTAG.

During the 1st Earth Observation Long Term Data Preservation workshop in May 2008, the draft guidelines and the framework were presented and debated by all European and Canadian EO data owners, data providers and archive holders. The participants discussed and developed a joint strategy to move ahead technically and programmatically concerning the Long Term Data Preservation of EO Data and recognized the need and benefits of a

common approach. Furthermore all the participants identified and agreed the draft LTDP Common Guidelines presented at the workshop as a first concrete and fundamental step to move ahead in creating the Long Term Data Preservation Framework. The guidelines should be adopted for old missions with a step-wise approach and straightforward for new missions and projects. ESA was given the task to trigger and coordinate the following steps toward the progressive European LTDP Framework implementation.

A consolidated European LTDP Common Guidelines document has been produced by the LTDP working group on the basis of the comments and feed-backs received during the LTDP.

## 4.13 ETH Data Repository

### 4.13.1. Introduction

The research of the Institute of Cartography is focused on topographic cartography (relief representation), thematic cartography, and atlas cartography (school atlases, national atlases). These domains are still considered the main fields of application, but adapted to new demands, media and technologies. The main aim in research is the development of new cartographic information and interaction methodologies.

The information in this section is compiled from the ETH (Eidgenössische Technische Hochschule) website (http://www.karto.ethz.ch/) and the questionnaire returned by Lorenz HURNI (ETH Zurich).

### 4.13.2. Archive systems

The currently used system is based on file server and RAID.

### 4.13.3. Archive access and connection

Above all it is an internal archive with no access for external users. Special products are accessible for users with password. The services support OGC WMS.

### 4.13.4. Archive content

The repository contains nearly 1 TB topographic map data of Switzerland based on aerial acquisitions. The data grown is estimated by 300 GB. The products keep OpenGIS WMS/ISO:19128 2005.

### 4.13.5. Data curation and long-term preservation

Actual data are sustained by backup copy and a recovery system. At present there is no commission for a long-term preservation. So data sets are overwritten by newest data.

## 4.14 EUMETSAT Data Repository

The main purpose of the European Organisation for the Exploitation of Meteorological Satellites (EUMETSAT) is to deliver weather and climate-related satellite data, images and products 24 hours a day, 365 days a year. This information is supplied to the National Meteorological Services of the organisation's member and cooperating states in Europe, as well as other users world-wide. EUMETSAT is an intergovernmental organisation and was founded in 1986. The information in this section is compiled from: www.eumetsat.int and the questionnaire returned by Sally Wannop (EUMETSAT).

### 4.14.1. Archive systems

The EUMETSAT data repository is located at EUMETSAT, Am Kavalleriesand 31, D-64295 Darmstadt, Germany. The EUMETSAT archive system is the Unified Meteorological Archive and Retrieval Facility (U-MARF), see Figure 22 for an illustration of U-MARF and its components. It is on Sun cluster environment running Solaris 10, with Sun StorageTek (SL8500 & T10000) and LTO-4 drives and on-line RAID array.

It runs with back-up copy, as a redundant and recovery system. Media store copy is on LTO-4 tapes. Data source are satellites, data subject is atmosphere/weather, land surface and water.

The actual amount of data spans 350 TB (approx 16 million files). The expected amount of GB growth per year is currently 95 TB/ year. In future (from about 2018 onwards), it is estimated to be 3000 PB/year.

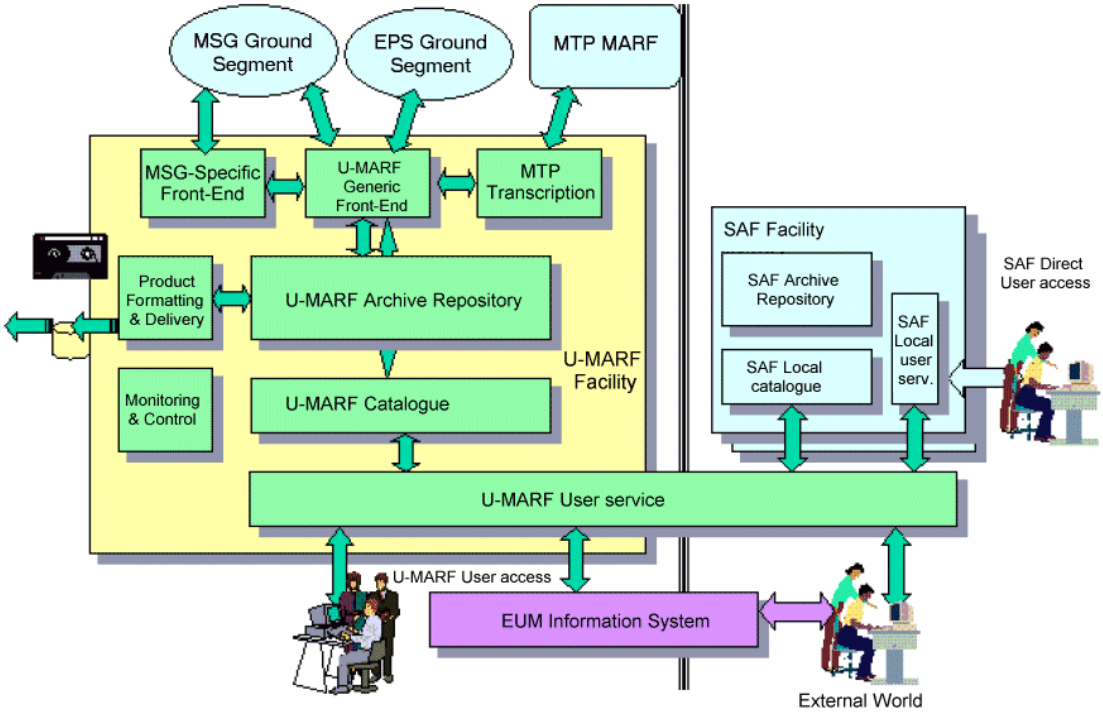


**Figure 22 - EUMETSAT U-MARF Archive Repository**

### 4.14.2. Archive access and connection

Subject to an annual review process, EUMETSAT archived data will be supplied free of charge. The full policy statement is explained in their Archive User Guide.

Furthermore, customers are requested to indicate their acceptance of the terms and conditions under which the data are supplied. All EUMETSAT satellite images are subject to EUMETSAT copyright. If one wishes to re-use any images, EUMETSAT's copyright credit should be shown by displaying the words "copyright © {year} EUMETSAT" on each of the images shown, where {year} is the current year.

The EUMETSAT Product Navigator is the central data discovery service for all EUMETSAT data and products including third-party products disseminated via EUMETCast. Users can search for data, order data via the EUMETSAT Archive, and subscribe to EUMETCast disseminated services.

The Product Navigator collection entries are compatible with ISO 19115/19139 standards and conform to the EU INSPIRE directive.

The EUMETSAT Product Navigator can be used to search and obtain details about the following resources from EUMETSAT:

- *Dataset:* for the satellite products;
- *Document:* for supporting documents, e.g. product guides, validation reports;
- *Software:* for software available from EUMETSAT.

Each product is presented with a brief description and a range of important information such as coverage, how the product is disseminated, for whom it is available, what the typical file formats are, examples of file naming conventions, and dissemination frequency. Links are provided to where the product can be accessed as well as to more information on the product itself.

In addition, an interactive map can be used to define a spatial area to search for data and products available within a specific geographical region.

The data, products and services provided by EUMETSAT are made available to users via the following key delivery methods:

- **EUMETCast** - established by EUMETSAT as the prime dissemination mechanism for Meteosat and Metop image data and meteorological products, EUMETCast is also used to deliver data supplied by several external data providers. Using similar technology to digital satellite TV, EUMETCast delivers meteorological and environmental data services to a wide audience. Reception station equipment is based on industry standard components. View the EUMETCast Dissemination Schedule.
- **Direct Dissemination** - is the traditional way to receive the image data and products directly from the satellite that acquires the atmospheric/earth observation data. To receive these data, users need dedicated reception station equipment.
- **Global Telecommunications System** (GTS)/RMDCN - established by the World Meteorological Organization, the GTS is used by national meteorological services throughout the world to receive, send and relay meteorological data and products. The Regional Meteorological Data Communication Network (RMDCN) is used by WMO Region VI to carry the GTS traffic within Europe. EUMETSAT and its partners use the GTS/RMDCN to deliver the following services:
    - Dissemination of Metop Level 1 Data;
    - Data Collection Retransmission;
    - Meteorological and environmental products generated by EUMETSAT and the Satellite Application Facilities (SAFs);
    - Meteorological Data Dissemination;
    - EUMETSAT ATOVS Retransmission.
- **Internet** - negating the need for satellite reception station equipment, Internet delivery offers an affordable way to transfer smaller volumes of data and products. The following services are available via Internet delivery:
    - Archive (historical) image data and meteorological products;
    - Archive Direct;
    - Low Rate SEVIRI Image data;
    - Data Collection Retransmissions;
    - Derived Product Imagery.

- **EUMETSAT website** - samples of processed Meteosat image data and derived product images can be viewed directly from this website.

### 4.14.3. Archive content

EUMETSAT provides an extensive range of high-quality satellite data, products and services. The primary services of EUMETSAT are those associated with the data from Meteosat first and second generation satellite series, and the Metop satellite series. In addition, EUMETSAT plays an important role in the retransmission of a variety of other satellite data, meteorological products and environmental data generated by their partners and a growing network of external data providers.

Standard data formats are according to internal policy: WMO (BUFR, GRIB) Standards, HDF Standards. Metadata standards are according to internal policy ISO 19115:2003 ISO 19139. It is planned to be also WMO WIS and INSPIRE compliant in future.

Data are archived from following satellites (instruments): Meteosat (MVIRI, SEVIRI, GERB), Metop (IASI, AVHRR, HIRS, AMSU-A, MHS, ASCAT, GOME-2, GRAS), Jason (Altimeter), NOAA (AVHRR, HIRS, AMSU-A, MHS).

### 4.14.4. Data curation and long-term preservation

EUMETSAT sustainability strategy is that it operates with long-term programmes - e.g. current MSG (Meteosat Second Generation) programme is funded until end 2018. New future programmes inherit the existing archive - e.g. future MTG (Meteosat Third Generation) programme will cover geo-stationary data until about 2035.

The quality of archive is sustainable to maintain, though technology improvements will be required in 2016-2018 timeframe to meet the increased data rates. Main obstacles for sustainability are money and technology issues. There are ongoing projects to implement archive interoperability using HMA / OGC / WMO / INSPIRE standards.

Dedicated reprocessing campaigns are undertaken when needed (a dedicated reprocessing environment is available and is being upgraded). A product versioning policy is implemented to distinguish between different versions of used software.

Hardware upgrade is performed approximately every 5 years due to capacity, speed, obsolescence, maintenance cost and to cope with evolving performance requirements.

All format definitions (documentation) and reader software are maintained. All archived data is migrated approximately every 5 years - but maintaining the formats (i.e., no migration of formats).

Discard protection is by second copy stored off-site. A third copy is maintained on-line for the most requested data.

The effort of archive related work is 7 man-years / year. Curation (migration) is performed as a dedicated project with its own resources (is not a permanent activity). Costs are approximately 2,000 K€ per year (for manpower, licences, maintenance, consumables).

## 4.15 EURAC Data Repository

### 4.15.1. Introduction

The European Academy Bozen/Bolzano (EURAC) is an institution for applied research and further education. The Institute for Applied Remote Sensing develops remote sensing applications for the Alps and other mountainous areas. The projects and research activities are addressing the current most pressing environmental problems within the context of global change and with respect to implications at the local and regional level. The information in this section is compiled from:

http://www.eurac.edu/Org/alpineEnvironment/RemoteSensing/index.htm

and the questionnaire returned by Stefan Schneiderbauer (EURAC).

### 4.15.2. Archive systems

The archive for satellite data is in its initial stage. Only little information can be given now. The system is based on EMC². Data sets are stored on tape. The use of standards for services is also under construction.

### 4.15.3. Archive access and connection

By means of a web site portal selected users with password can download products incl. raw data. The archive has high speed connectivity via HiSEEN, GEANT.

### 4.15.4. Archive content

The archive will start storing AQUA Satellite data and Ground Based Sensor data from the local area. The data grown is estimated at 1 TB.

### 4.15.5. Data curation and long-term preservation

Data curation and long-term preservation is planned but under construction.

## 4.16 EuroGeographics Data Centre

EuroGeographics represents 50 National Mapping and Cadastral Agencies from 46 countries across Europe.

### 4.16.1. Archive systems

The Data Centre is located in Frankfurt (Germany), based on an ArcGIS Geodatabase with backup (three different locations in Europe) and recovery systems. Off-line copy is stored on DVD.

### 4.16.2. Archive access and connection

The data are available via the web site portal and the original raw data can be obtained on physical media. The products are licensed on an annual basis and with a minimum license period (from 2 to 5 years).

### 4.16.3. Archive content

EuroGeographics currently has 4 pan-European geographic datasets:

- EuroDEM: digital representation of the ground surface topography of Europe;
- EuroBoundaryMap: digital geographic reference database at the scale 1:100 000;
- EuroGlobalMap: digital topographic dataset covering Europe at the scale 1:1 Million;
- EuroRegionalMap: multi-functional topographic reference dataset at the scale 1:250 000.

The total amount of data is 1 TB, in ESRI Shapefile format with metadata compliant with the ISO 19115:2003 and ISO 19139 standards.

### 4.16.4. Data curation and long-term preservation

The long-term interpretability is secured with compliance with the INSPIRE policy.

Data reprocessing is subject to a licensing policy.

## 4.17 Euromap Data Repository

### 4.17.1. Introduction

Euromap GmbH is a company within the business of Earth observation, located in Germany. Euromap represents the only IRS (Indian Remote Sensing Satellite) receiving and distribution facilities in Europe. Euromap provided data to the GMES FTSL project via the prime contractor.

(source: http://www.euromap.de/)

### 4.17.2. Archive systems

Internal systems of 100 GB storage.

### 4.17.3. Archive access and connection

Data are ordered via customer service and licenses to products can be purchased.

No raw data are given to external users, no access via gateway, no high speed network.

IRS data products are offered on the following media:

- CD-R (default), ISO 9660 file system with Joilet and Rockridge extensions, 650 MB;
- DVD-R, ISO 9660 file system with Joilet and Rockridge extensions, 4.7 GB;
- USB Hard Disk, external hard disk drive using FAT32 file system;
- Online, via Internet using the FTP protocol.

### 4.17.4. Archive content

IRS data products have been used successfully in the fields of agriculture, forestry, geology and hydrology, for mapping and planning tasks and for other applications where map scales of up to 1:25'000 were required.

The information provided at the webpage is valid for data from the European footprint. Data from other ground stations might be priced differently and may not be available in all listed product sizes, processing levels and formats.

### 4.17.5. Data curation and long-term preservation

On the website the data formats used for the products are mentioned and an associated document provides the description of the format: Fast Format, Super Structure Format, GeoTiff, RPC file, and metadata file. Additional value added applications are provided such as a 3D Terrain Visualisation and maps such as 5 m natural colour ortho-mosaics on financial conditions.

## 4.18 EUSC Data Repository

### 4.18.1. Introduction

The mission of the European Union Satellite Centre (EUSC) is to support the decision-making of the European Union by providing analysis of satellite imagery and collateral data.

The EUSC is an Agency of the Council of the European Union. It is one of the key institutions for European Union's Security and Defence policy, and the only one in the field of

space (see http://www.eusc.org/). The Centre shall, in coherence with the European Security Strategy, support the decision making of the European Union in the field of the Common Foreign and Security Policy (CFSP), in particular of the European Security and Defence Policy (ESDP), including European Union crisis management operations, by providing, as appropriate, products resulting from the analysis of satellite imagery and collateral data, including aerial imagery, and related services. The European Union Satellite Centre is located in Torrejón de Ardoz, in the vicinity of Madrid, Spain.

### 4.18.2. Archive systems

The repository is on a WINDOWS/ORACLE platform. Media store copy is on DVD/LTO. The amount of data is about 12 TB.

### 4.18.3. Archive access and connection

No access to external users.

### 4.18.4. Archive content

The repository contains satellite data from SPOT, IKONOS, TERRASAR, EROS, RADARSAT, LANDSAT and QUICKBIRD related to security.

### 4.18.5. Data curation and long-term preservation

No information could be obtained on data curation and long-term preservation.

## 4.19 ICCM Data Repository

### 4.19.1. Introduction

The Instituto Canario de Ciencias Marinas (ICCM) is located on Gran Canaria (Canary Islands). It is a governmental institute devoted to marine sciences. It maintains a data repository containing data relevant for the ocean sciences. The information below is extracted from the questionnaire returned by Javier Perez-Marrero (ICCM).

### 4.19.2. Archive systems

Data repository is on hard disks within the organisation, back-up copy exists.

### 4.19.3. Archive access and connection

Access is given to registered external users with appropriate tools. Raw data are not distributed to external users.

### 4.19.4. Archive content

Data are from satellite and other database for the central North Atlanic: AVHRR, SEVIRI, SeaWIFS, MODIS, Quikscat. Archive contains approximately 400 GB and is increasing about 25 GB/year. Data format and metadata format are according to internal policy.

### 4.19.5. Data curation and long-term preservation

The curation is organized to verify and keeping track of volatile information. Main obstacle for sustainability is manpower uncertainty.

## 4.20 INFOTERRA Data Repository

### 4.20.1. Introduction

Infoterra is a commercial provider of geo-spatial products & services. For example, Infoterra supports applications such as, flood risk analysis, network planning, humanitarian relief and geological mapping (see http://www.infoterra.co.uk/ for more information). The information below has been provided by Thomas Lankester (Infoterra).

Infoterra hosts 4 data archives:

- The UK-MM-PAF;
- The Respond map catalogue;
- The Infoterra processed MERIS Level 3 products;
- The Bluesky/Infoterra GeoPerspectives aerial photography and DEM collections.

The UK-MM-PAF is operated on behalf of ESA. Secondly, as part of the RESPOND project, Infoterra maintains the project map catalogue. Thirdly, Infoterra has generated an archive of Level 3 composite products based on MERIS Reduce Resolution data from 2002 to the present. Fourthly, the Bluesky/Infoterra GeoPerspectives airborne Data contains an extensive range of aerial survey geo-information products (aerial photography, colour infrared imagery, digital terrain and digital surface models) of England & Wales, and lidar height data for most British cities.

### 4.20.2. Archive systems

#### 4.20.2.1. UK-MM-PAF Archive systems

The UK-MM-PAF uses the Multi-Mission Facility Infrastructure (MMFI) to process, archive and disseminate ERS and ENVISAT data products. The archive element of MMFI is the Local Inventory which is composed of the data catalogue and an Archive Management System. The archive type is near-line. The estimated volume is 260 TB, with an increase of 3 TB per month.

#### 4.20.2.2. Respond Archive system

Apache Web server running on a Linux Internet server. Archive type is on-line. The estimated volume is 5 GB, with an increase of 25-75 MB per month.

#### 4.20.2.3. Level 3 archive system

Apache Web server running on a Linux Internet server. MapServer Web Map Service running on a Linux Internet server. Archive type is on-line with an LTO3 tape back-up (not digitally indexed/catalogued). The estimated volume is 300 GB, with an increase of 6 GB per month.

#### 4.20.2.4. Bluesky/Infoterra GeoPerspectives archive systems

JBOSS and Apache Web server running on a Linux Internet server. Archive type is on-line with off-line back-up. The estimated volume is 38 TB, with an increase of 500 GB per month.

### 4.20.3. Archive access and connection

#### 4.20.3.1. UK-MM-PAF Archive access and connection

Archive access is via ESA Product Order Handling (POH). Both on request and on-line distribution of data are supported. The ESA Data Policy applies. Delivery is possible on CD-ROM, DVD, FTP pick-up or via the MERCI on-line system (for the (A)ATSR dataset series ATS_TOA_1P, ATS_AR__2P, ATS_NR__2P). Off-line processing at UK-MM-PAF is already implemented.

#### 4.20.3.2. Respond archive access and connection

The service offers on-line Web pages, map search, GeoRSS feeds and download.

#### 4.20.3.3. Level 3 archive access and connection

The service offers on-line Web pages, browse (quicklook) images, FTP data download and an evaluation service conforming to the OGC WMS EO product extension.

#### 4.20.3.4. Bluesky/Infoterra GeoPerspectives archive access and connection

The archive is available via a Web server with registered user name and password access to data on-line via an OGC Web Map Service. Products can also be ordered for delivery via CD, DVD, FTP or hard disk.

### 4.20.4. Archive content

#### 4.20.4.1. UK-MM-PAF Archive content

The UK-MM-PAF hosts the following collections:

- ENVISAT / ASAR ASAR Level 0;
- ENVISAT / ASAR Alternating Polarisation Mode (X polar H);
- ENVISAT / ASAR Alternating Polarisation Mode (X polar V);
- ENVISAT / ASAR Imaging Mode;
- ENVISAT / ASAR Wide Swathe;
- ENVISAT / ASAR Global Monitoring Image Mode;
- ENVISAT / AATSR AATSR Level 0 Product;
- ENVISAT / AATSR AATSR Top Of Atmosphere gridded brightness temperature and reflectance;
- ENVISAT / AATSR AATSR averaged geophysical product;
- ENVISAT / AATSR AATSR geophysical product (full resolution);
- ESR-1 / SAR ERS SAR Level 0;
- ERS-2 / SAR ERS SAR Level 0;
- ERS-1 / Altimeter, ATSR ERS Low Bit Rate Level 0;
- ERS-1/Altimeter ERS Altimeter L2;
- ERS-2/Altimeter ERS Altimeter L2;
- ERS-1 / ATSR ERS-1 ATSR Level 1B (Envisat Format);
- ERS-1 / ATSR ERS-1 ATSR Level 2 (Envisat Format);

- ERS-1 / ATSR ERS-1 ATSR Level 2 (Envisat Format);
- ERS-2 / ATSR-2 ERS-2 ATSR-2 Level 1B (Envisat Format;
- ERS-2 / ATSR-2 ERS-2 ATSR-2 Level 2 (Envisat Format);
- ERS-2 / ATSR-2 ERS-2 ATSR-2 Level 2 (Envisat Format);
- ENVISAT / MERIS MERIS Full Resolution Level 0;
- ENVISAT / MERIS MERIS Full Resolution Full Swath Level 1;
- ENVISAT / MERIS MERIS Full Resolution Full Swath Level 2.

#### 4.20.4.2. Respond archive content

Raster data maps in JPEG format and GIS layers in KML.

#### 4.20.4.3. Level 3 archive content

The archive contains Level 3 data with composing periods of (variously) 8 days, 10 days and calendar months. The following collections are hosted in this archive:

- Case 1 water chlorophyll concentration: globally binned;
- Case 1 water chlorophyll concentration: regional flux conserving resampling;
- MERIS Terrestrial Chlorophyll Index: globally binned;
- MERIS Terrestrial Chlorophyll Index: regional flux conserving resampling.

#### 4.20.4.4. Bluesky/Infoterra GeoPerspectives archive content

The following collections are hosted in this archive:

- 25 cm sampled aerial photography of England and Wales;
- 5m sampled DTM (Digital Terrain Modeling) and DSM (Digital Surface Model) of England and Wales;
- UK airborne LIDAR data, principally of city centres, with 1m x-y sampling.

### 4.20.5. Data curation and long-term preservation

#### 4.20.5.1. UK-MM-PAF Data curation and long-term preservation

The data curation and long-term preservation is done by ESA, using the ESA procedure (see 4.12).

#### 4.20.5.2. Respond data curation and long-term preservation

Basic map metadata (e.g. location, producer, date) is maintained along with the maps themselves (see http://www.respond-int.org/respondlive/searchmaps.html). Long-term preservation will be migrated to the SAFER GMES FP7 project.

#### 4.20.5.3. Level 3 data curation and long-term preservation

Metadata is stored in an XML file with a unique Data Type Definition (DTD).

Long-term this metadata will be converted to an XML format compliant to the ISO 19139 schema.

#### 4.20.5.4. Bluesky/Infoterra GeoPerspectives data curation and long-term preservation

All data is preserved on the original delivery media in addition to being loaded onto the SAN (Storage Area Network) where it is automatically backed up to tape storage via the hierarchical storage management policy. Data can be restored by location and should be considered as a data backup rather than an archive.

## 4.21 INGV Data Repository

### 4.21.1. Introduction

INGV was meant to gather all scientific and technical institutions operating in Geophysics and Volcanology and to create a permanent scientific forum in the Earth Sciences. INGV cooperates with universities and other national public and private institutions, as well as with many research agencies worldwide. The new institution, currently the largest European body dealing with research in Geophysics and Volcanology, has its headquarters in Rome and important facilities in Milano, Bologna, Pisa, Napoli, Catania and Palermo.

The main mission of INGV is the monitoring of geophysical phenomena in both the solid and fluid components of the Earth. INGV is devoted to 24-hour countrywide seismic surveillance, real-time volcanic monitoring, early warning and forecast activities. State-of-the-art networks of geophysical sensors deliver a continuous flow of observations to the acquisition centres of Rome, Naples and Catania, were the data are analyzed around the clock by specialized personnel. In addition to being analysed for research and civil defence purposes, the data supplied by numerous monitoring networks are regularly distributed to the public institutions concerned, to the scientific community and to the public.

INGV operates in close coordination with the Ministry of University and Research and with Civil Protection authorities, both at national and local level. INGV also cooperates with the Ministry of Environment, the Ministry of Education, the Ministry of Defence and the Ministry of Foreign Affairs in the frame of large research programs of national and international relevance.

INGV pays special attention to Education and Outreach through publications for schools, scientific exhibitions and dedicated Internet pages.

The information in this section is compiled from http://www.eurac.edu/Org/alpineEnvironment/RemoteSensing/index.htm the INGV web site (http://www.ingv.it) and the questionnaire returned by Maria Fabrizia Buongiorno.

### 4.21.2. Archive systems

Data archiving is managed by a server and storage is made on RAID disks.

### 4.21.3. Archive access and connection

No information has been given in the questionnaire. A lot of information is accessible freely on the website.

### 4.21.4. Archive content

The archive contents NOAA-AVHRR data acquired by the INGV receiving antenna.

### 4.21.5. Data curation and long-term preservation

There is no redundant system nor recovery system. Preservation is made by back-up copy.

# 4.22 INSA Data Repository

## 4.22.1. Introduction

INSA, created in 1992, is the largest Spanish company in the space sector by number of employees and within the field of space services INSA is a commercial firm 100% property of the National Institute for Aerospace Technique (INTA), a body of the Spanish Ministry of Defence.

INSA's activities are divided in two main areas:

- Ground Stations: INSA is responsible for the operations and maintenance of satellite tracking and data reception facilities in NASA's (Robledo de Chavela/Madrid), ESA's (Villafranca/Madrid and Cebreros/Ávila) and INTA's (Maspalomas/Gran Canaria) space centres in Spain. INSA also performs Quality Assurance activities in French Guiana (Kourou).
- Engineering: INSA's engineering activities comprise all the phases from definition to customer support after delivery of ground segment systems, satellite communications, and new services based on aerospace systems (in particular risk management systems).

The information in this section is compiled from the INSA web site (http://www.insa.es) and the questionnaire returned by Susana Martinez.

See also the RISK-EOS chapter.

## 4.22.2. Archive systems

The archive system is composed of PC, external hard disks and DVD.

## 4.22.3. Archive access and connection

The Data Repository is not connected to public network. There is an internal access via a secured (password) connection.

## 4.22.4. Archive content

Data is providers are Landsat, Spot, MSG, TERRA and Aqua. The amount of data is 80 GB.

## 4.22.5. Data curation and long-term preservation

Data curation is based on internal policy with INSPIRE compliant metadata.

The preservation is based on back-up copies and a redundant system.

# 4.23 INTEGRAL Data Repository

## 4.23.1. Introduction

INTEGRAL (Interferometric Evaluation of Glacier Rheology and Alterations) is an initiative of the Sixth Framework Programme of the European Commission coordinated by the Joanneum Research Institute (see Website: http://dib.joanneum.at/integral). Its general objective is to promote an advanced observation technology for the unsupervised detection, precise measurement and variational analysis of ice motion / deformation on large European glaciers based on the complementary use of radar interferometry and interferometric altimetry, and to support natural exploration, social-economic activities and subsequent surveys in the glacier environment with equivalent Theological models and appropriate information on the glacier regime in the form of new value-added INSAR products.

### 4.23.2. Archive systems

The INTEGRAL data repository is located at the Institute of Digital Image Processing of the Joanneum Research in Graz (Austria) on a Windows 2003 FTP-Server (Intel P4 1,6 GHz, 2 GB RAM, Storage: 100 GB). It disposes of backup and recovery systems. The off-line copy of the data is stored in LTO Ultrium 3 Backup Tapes (800 GB).

### 4.23.3. Archive access and connection

The repository is accessible with appropriate tools by selected users with password. It is connected to a high-speed network (ACOnet) and accessible via a gateway.

It is not possible to obtain the original raw data and any process capacity is provided to external users.

### 4.23.4. Archive content

The data are new value-added INSAR products based on SAR data from different satellites (ERS, ENVISAT, ICESat, JERS, ALOS, NOAA, ASTER, ...), covering the entire Europe since 1980s.

The total amount of data is 10 GB, with an expected increase of 5 GB/year.

### 4.23.5. Data curation and long-term preservation

The 10% of the total archive management effort is spent in data curation, following an internal policy (no documentation available). The long term access will be assured by a data-exchange plan and the migration to OpenSource products.

## 4.24 JRC Data Repositories

### 4.24.1. Introduction

The CID (Community Image Data portal) is an Action launched under the 7th Framework Program, in the MARS Unit. It relates to the acquisition, access and storage of Satellite (and aerial) Remote Sensing data in the context of the JRC and Commission Services' needs.

The conception, population, maintenance and dissemination of the satellite remote sensing data archive, while respecting the applicable Intellectual Property Rights, are implemented in the CID Image Data Portal by the CID action.

The JRC Data Repository contains datasets that have been collected between 1990 and 2009. For the time being the majority of the data has been collected for control purposes in the context of the EU Common Agriculture Policy. The so-called Control with Remote Sensing programme has been set up in close collaboration between JRC MARS Unit and DG Agriculture. This programme is currently one of the biggest European civil projects using satellite remote sensing data.

Starting in summer/autumn 2009 the CID repository will also include Time-series of Low/Medium Resolution data used for agricultural yield estimates: vegetation index and agromet products (like NDVI, fAPAR, SAVI) based on AVHRR, Spot VGT, Modis, MSG. Geographic extent are a Pan-European coverage 1981-2009 and 10 regions of interest for global food security monitoring.

The CID Portal is a Web portal to search and access Remote Sensing data and derived products hosted at JRC. Search and preview is open to the public. Access to the image data via a variety of online services is limited to authorized and registered users. Staff of institutions of the European Union can be granted access to most of the archive data after registration.

### 4.24.2. Archive systems

The archive is physically located at the JRC in Ispra. Data storage is made of NetApp Network Attached Storage (NAS) appliance provided file protocols: NFSv4 (Network File System v4) and CIFS (Common Internet File System).

### 4.24.3. Archive access and connection

The first release of the CID portal became public at the end of August 2008. The main entry point and discovery service for users who want to search for images in the catalogue is the CID Portal Web application (http://cidportal.jrc.ec.europa.eu/imagearchive). The interface which is available to the public, allows searching the archive for images based on various filter parameters, such as platform, acquisition date, image resolution, geo-graphic area, etc. Data matching the search criteria are listed with their metadata and can be previewed via image thumbnails.

Since large parts of datasets are covered by restrictive IPR's, any further in-depth access to the datasets requires authorization and hence a user registration. After having logged in and depending on their credentials and privileges, authorized users can view the images or load them into various applications via the HTTP based protocol ECWP. They can define so-called portfolios for the data access services FTP, WMS and WCS. Users inside the EC JRC network are also able to access the data via standard file system mechanisms.

The CID repository is connected to the Internet via HISEEN/GEANT network.

### 4.24.4. Archive content

The geographical coverage of the archived High and Very High Resolution data is mainly the area of the EU Member States. Data resolution (ground sampling distance) ranges between 0.6 and 30 m.

The archive content is:

- SPOT 1-5 High and Very High Resolution data: 1.3 TB;
- Landsat: High Resolution data: 300 GB;
- IRS: High Resolution data: 200 GB;
- Radarsat: 150 GB;
- Quickbird: Very High Resolution data: 4 TB;
- Ikonos: Very High Resolution data: 8 TB.

### 4.24.5. Data curation and long-term preservation

Data is stored on-line on highly redundant disk RAID systems. Short-term backup is performed through volume snapshots. Long-term backup is ensured by regular backup on tape with indeterminate safekeeping. Regular file checksum tests are planned to be implemented to identify problems with file system consistency.

## 4.25 KSAT Data Repositories

### 4.25.1. Introduction

Kongsberg Satellite Services AS (KSAT) is a commercial Norwegian enterprise, providing services based on data from polar orbiting satellites such as:

- Telemetry, Tracking and Command services (TT&C);

- Global data dump services;
- Operational Earth Observation.

The company currently operates three ground stations: the Tromsø Station at 69°39´N 18°56´E, Svalbard Satellite Station (SvalSat) at 78°15´N 15°80´E, Grimstad, (South Norway) at 58°20´N 8°21´E and TrollSat 72°S 2°E.

The Northern locations of Tromsø and SvalSat are ideal for data acquisition and TT&C services from polar orbiting satellites. KSAT is in the unique position to offer operational services for all polar orbiting passes due to the exclusive latitude of the ground stations. As TrollSat became operational spring 2007, the Southern location enables KSAT to utilize the concept of Pole-to-Pole interaction, with the ability to dump data and perform TT&C on both poles.

The head office of KSAT is located in Tromsø, and employs today 52 of the 67 employees at KSAT. The remaining 15 are located at the SvalSat ground station at Svalbard.

KSAT is operational around the clock all year.

KSAT is built on the former Tromsø Satellite Station (TSS) organization with history in the Earth Observation business that goes back to the 1960's. The Norwegian Space Centre and Kongsberg Defence & Aerospace own KSAT on a 50/50 basis.

The information in this section is collected from the questionnaire returned by Mai-Eli Johansen and from the KSAT web site.

### 4.25.2. Archive systems

The KSAT - internal library is based on Sun StorageTek L700 tape library.

The MMFI (Multi Mission Facility Infrastructure) consists of a StorageTek robot from SUN and a number of different computers. The MMFI is a system for archiving, retrieving and ordering of different kinds of satellite data and is owned and maintained by ESA.

### 4.25.3. Archive access and connection

The archive access is restricted to KSAT registered customers via a KSAT GUI.

### 4.25.4. Archive content

The KSAT - internal library contains RSAT-2 data from June 2009. RSAT-1 data is planned to be archived and it has a covers the period from 1995 until now. RSAT-2 on line data is a total of 1.3 TB. Of line RSAT-1 data planned to be stored in the repository is approximately a total of 80 TB.

MMFI holds Kompsat-2 level 1r and level 1g optical images. The data covers a period from June 2008 and has a volume of 900 GB.

### 4.25.5. Data curation and long-term preservation

KSAT archives all data in a off-line archive. The environment in the off-line archive is controlled regarding temperature and humidity. The archive is not fire safe and we do not have a backup of the tapes in a separate building. They do not yet have any policy regarding copying of data onto new media when the tapes get old and more fragile.

## 4.26 MRI Data Repository

### 4.26.1. Introduction

The Ministry for Resources and Infrastructure (MRI) (www.mri.gov.mt) was established in 2002 and is mainly responsible for major infrastructural works and projects that are carried out throughout the Maltese Islands. The main operational organisation within the Ministry is the Works Division which works in close conjunction with the Building Industry Consultative Council (BICC), as the official link between the private sector and Government on matters related to the construction industry, and the Malta Resources Authority (MRA)." (as mentioned in GINIE: Geographical Information Network in Europe, IST-2000-29493, Survey of key GI players within Europe, GINIE D2.1.1, Author: Mark Probert, Date 11/11/2003).

The information in this section is compiled from http://www.mri.gov.mt/ and the questionnaire returned by Victor Vella (National Agriculture agency).

### 4.26.2. Archive systems

Up to now no regular archiving is done. Data sets are stored on DVD.

### 4.26.3. Archive access and connection

The data sets are only for internal use.

### 4.26.4. Archive content

The archive contains 40 GB QuickBird 2 scenes of Europe with relevance for Malta.

The data growth is estimated at 10 GB.

### 4.26.5. Data curation and long-term preservation

Up to now no data curation and long-term preservation is planned.

## 4.27 NERSC Data Repository

### 4.27.1. Introduction

Nansen Environmental and Remote Sensing Center (NERSC) is an independent non-profit research institute affiliated with the University of Bergen, Norway. The information in this section is compiled from http://www.nersc.no and the questionnaire returned by Stein Sandven (NERSC).

### 4.27.2. Archive systems

The actual archive is based on LINUX servers and file archive without database.

### 4.27.3. Archive access and connection

NERSC provides a web side portal with OGC service WMS. External user can access to products after authorisation for free. Raw data are not accessible.

### 4.27.4. Archive content

The archive contains SAR ocean products, Ocean colour products and Sea ice from multisensor data in EO format (CEOS, DIMAP...). It stores mainly retrieved products. Raw data are only stored if it is difficult to recover from original archives. Metadata are INSPIRE compliant. The archive covers a time frame from 2005 (partly from 1978) to present.

### 4.27.5. Data curation and long-term preservation

Up to now NERSC pursues no comprehensive strategy for Data curation and long-term preservation.

## 4.28 NILU Data Repositories

### 4.28.1. Introduction

The Norwegian Institute for Air Research (NILU) conducts environmental research with emphasis on the sources of airborne pollution, atmospheric transport, transformation and deposition and is also involved in the assessment of the effects of pollution on ecosystems, human health and materials. A main priority for NILU is to provide scientific facts on the quantitative relationships between these factors, and at the same time make the results available in user friendly manners for decision-makers. NILU hosts an ESA sponsored database (http://nadir.nilu.no/calval/) used for archiving of campaign data. It is used extensively for Envisat correlative data and Cal/Val of AATSR, MERIS, SCHIMACHY, MIPAS and GOMOS. Later on generalised and expanded to allow storage of other data types such as hyper spectral and SAR images. NILU also operates an archive for Groundbased in-situ data (http://ebas.nilu.no/) covering atmospheric composition data mainly from rural European sites. Focus on log-range transport of air pollution. The information in this section is compiled from web pages mentioned above and the questionnaire returned by Aasmund Fahre Vik (NILU).

### 4.28.2. Archive systems

The Calval file archive stored at backed-up server. Files are indexed and searchable through database-driven web interface or via simple ftp.

EBAS data are stored in a SYBASE relational database running on a SUN Solaris server and data are made available through .net web application running on a Red Hat Linux server. tape

### 4.28.3. Archive access and connection

For Calval data an authentication and authorisation is needed. Users may download whatever data they access to. The archive supports http or ftp and the provision of off-line media CD-ROM or DVD.

Also for EBAS a partial Authentication / Authorisation is needed.

The archives have High speed internet connectivity via HISEEN/GEANT.

### 4.28.4. Archive content

The Calval archive contains satellite data from ENVISAT Cal/Val, Chris/Proba, ESAR, etc. but mainly Level 2 atmospheric composition measurements from non-satellite platforms. Data sets archived since 2001 and onwards. Till now the capacity is nearly 100 GB with a yearly growth of ~ 5 GB.

EBAS contains European Ground Based Sensor data from 1970 to now. The capacity is less than 100 GB.

### 4.28.5. Data curation and long-term preservation

The archives use backup and recovery systems.

Calval data uses HDF4/5 format and CEOS-WGCV-ACSG metadata standards (compatible with AURA validation data centre). Images stored as they are, but metadata must either be provided manually through web interface or stored in an HDF-file (which the image then should be attached to).

## 4.29 NINA Data Repository

### 4.29.1. Introduction

The Norwegian Institute for Nature Research (NINA) is Norway's leading institution for applied ecological research. NINA is responsible for long-term strategic research and commissioned applied research to facilitate the implementation of international conventions, decision-support systems and management tools, as well as to enhance public awareness and promote conflict resolution. The information in this section is compiled from http://www4.nina.no/ninaeng/ and the questionnaire returned by Frank Hanssen (NINA).

### 4.29.2. Archive systems

Data sets are stored on disk.

### 4.29.3. Archive access and connection

Nina provides common OGC servicesbut no access for external users.

### 4.29.4. Archive content

The archive contains LandSat and Aerial data sets in OpengGIS Web Map Context ESRI Shapefile data format.

### 4.29.5. Data curation and long-term preservation

Up to now beside backup copies NINA pursues no comprehensive strategy for Data curation and long-term preservation.

## 4.30 NLS Data Repository

### 4.30.1. Introduction

National Land Survey (NLS) of Finland produces and provides information on and services in real estate, topography and the environment for the needs of citizens, other customers and the community at large. The National Land Survey (NLS) is responsible for Finland's cadastral system and general mapping assignments. It also promotes the shared use of geographic information.

The NLS consists of 13 District Survey Offices, five national operational units and the small central administration. The NLS has staff of over 2000, of whom over 80 % are employed in the District Survey Offices. The NLS is a governmental agency subordinate to the Ministry of Agriculture and Forestry.

The information in this section is compiled from the NLS website (http://www.maanmittauslaitos.fi) and the questionnaire returned by Antti Saarikoski.

### 4.30.2. Archive systems

The archive is based on a WMS service for raster data (topographic maps, ortophotos) from the territory of Finland based on Bentley software.

### 4.30.3. Archive access and connection

NLS topographic maps covering all Finnish regions are available at Citizen's MapSite. This service is universally accessible and free of charge. Citizen MapSite allows requests only from ".fi" domain.

Professional MapSite shows NLS topographic maps, cadastral data and sale price information. The service is subject to charge and requires a user contract. This service is only available in Finnish and Swedish.

### 4.30.4. Archive content

Archive contents are aerial photographs of Finland.

### 4.30.5. Data curation and long-term preservation

The preservation is made with back-up copies, redundant and recovery systems.

Metadata follows the ISO 19115-2 standard.

## 4.31 NOA Data Repository

### 4.31.1. Introduction

The NOA (National Observatory of Athens) brings together 5 institutes :

- Institute of Astronomy and Astrophysics;
- Institute of Geodynamics;
- Institute for Environmental Research and Sustainable Development;
- Institute for Space Applications and Remote Sensing;
- Institute for Astroparticles Physics, NESTOR.

The information in this section is compiled from the NOA website (http://www.noa.gr) and the questionnaire returned by Athanassios Ganas.

### 4.31.2. Archive systems

The archive is managed by Leica Spider NET V2.2.

### 4.31.3. Archive access and connection

Links to RINEX daily data from the Ionian Sea GPS network can be found on the following page : http://www.gein.noa.gr/gps.html.

NOA supports an Open Data Policy. Data are provided to the Science community under the agreement that proper acknowledgements are made.

### 4.31.4. Archive content

The data source is satellites.

The data is in Compact RINEX format.

The data covers a period from 2006 up to now.

The volume of data is about 250Gb.

### 4.31.5. Data curation and long-term preservation

There is no current policy for data curation and long-term preservation.

A back-up is made every 2 months. An upgrade of the hardware is made every 3 years and one of the software is made every year.

## 4.32 OGS Data Repositories

### 4.32.1. Introduction

The National Institute of Oceanography and Experimental Geophysics - OGS is a public-sector scientific institution which promotes, co-ordinates and carries out research in the Earth Sciences, with a particular emphasis on the development of new analytical tools and applied technologies in the marine sciences and in seismology, and in the mineral exploration and environmental management fields.

OGS is one of the Ocean Data Platform providing access to marine data, meta-data, products and services within SeaDataNet distributed infrastructure.

The information in this section is compiled from the OGS website (http://www.ogs.trieste.it) and the questionnaire returned by Alessandra Giorgetti.

### 4.32.2. Archive systems

The archive is on a Linux server managed by National Oceanographic Data Centre with Oracle database.

### 4.32.3. Archive access and connection

The archive is accessible through the website with a restricted access. The archive has high speed connectivity (GEANT, HiSEEN).

### 4.32.4. Archive content

The data source is ground based sensors. It concerns atmosphere, weather and water.

The covered period is from 1891 until now; it represents an amount of data of 7GB.

The data standard is based on an internal policy, MedAtlas, ODV and NetCDF. For metadata, they are based on an internal policy, ISO 19115:2003, ISO 19115-2, ISO 19139 and are INSPIRE compliant.

### 4.32.5. Data curation and long-term preservation

The long term preservation is based on a continuous upgrade of hardware and software.

## 4.33 Planetek Data Processing Centre

### 4.33.1. Introduction

Planetek Italia is one of the main operators in Italy in the field of information sciences applied to land management. The company is active in the development of information solutions for storage, elaboration and distribution of cartographic databases and satellite remotely sensed images. Planetek Italia is also involved in software development and resale and in consulting services and training. The information in this section is compiled from Planetek website (http://www.planetek.it) and the questionnaire returned by Daniela Drimaco (Planetek).

### 4.33.2. Archive systems

Planetek's archive system is based on RedHat Enterprise Linux ES Servers which manage data sets in a 1st stage: server storage backup (5x500 GB + 6x300 GB RAID 5) and a 2nd stage: tape backup (HP Ultrium LTO 200 GB).

### 4.33.3. Archive access and connection

The access is monitored with Enterprise Network Security Appliances to secure free access or/and selected user access with password.

### 4.33.4. Archive content

Planetek's archive contains 2.5 PB of Data sets aimed to Water, Land Surface and Forest monitoring. All data are stored in an internal archive to be used within different projects. Several data sources are used: Satellite, aerial, statistics, and ground based sensors.

Ikonos: from 1999 to selected temporal range. Spot: from 1996 to selected temporal range. Landsat: from 1996 to selected temporal range. MERIS, MODIS: from 2006 to selected temporal range. The yearly growth is expected in the range of 1000 GB/year.

### 4.33.5. Data curation and long-term preservation

Planetek manages the archive with backup, redundancies and recovery system for long-term preservation. It uses OGC standards in a wide range for content rendering, service discovery and description. The archive supports following data formats: OpenGIS GML OpengGIS Web Map Context ESRI Shapefile and its meta data are ISO 19115:2003 conform.

## 4.34 PREVIEW Data Repositories

### 4.34.1. Introduction

Telespazio, member of the Space Alliance between Finmeccanica and Thales, is a world leader in satellite services.

Inter alia Telespazio carries out all Earth Observation commercial activities: from acquiring and processing satellite data, to the development and sale of software and products. Furthermore, the company takes part of the ESA programme GMES, that will provide monitoring services for surveillance and homeland security. PREVIEW (PREVention Information and Early Warning) is an information services for assets mapping, risk mapping, risk monitoring, risk forecasting and awareness and damage assessment for of hazards.

The information in this section is compiled from Telespazio PREVIEW page (http://www.telespazio.it/gmeslprev.htm) and the questionnaire returned by Daniele Pellegrino (Telespazio).

### 4.34.2. Archive systems

PREVIEW data sets are partially stored on file system and partially saved on PostGis database. Storage media are hard disk and DVD.

The following Figure 23 presents PREVIEW archive logical architecture extracted from the following presentation:
http://earth.esa.int/workshops/2007Geohazards/participants/464/pres_464.pdf.

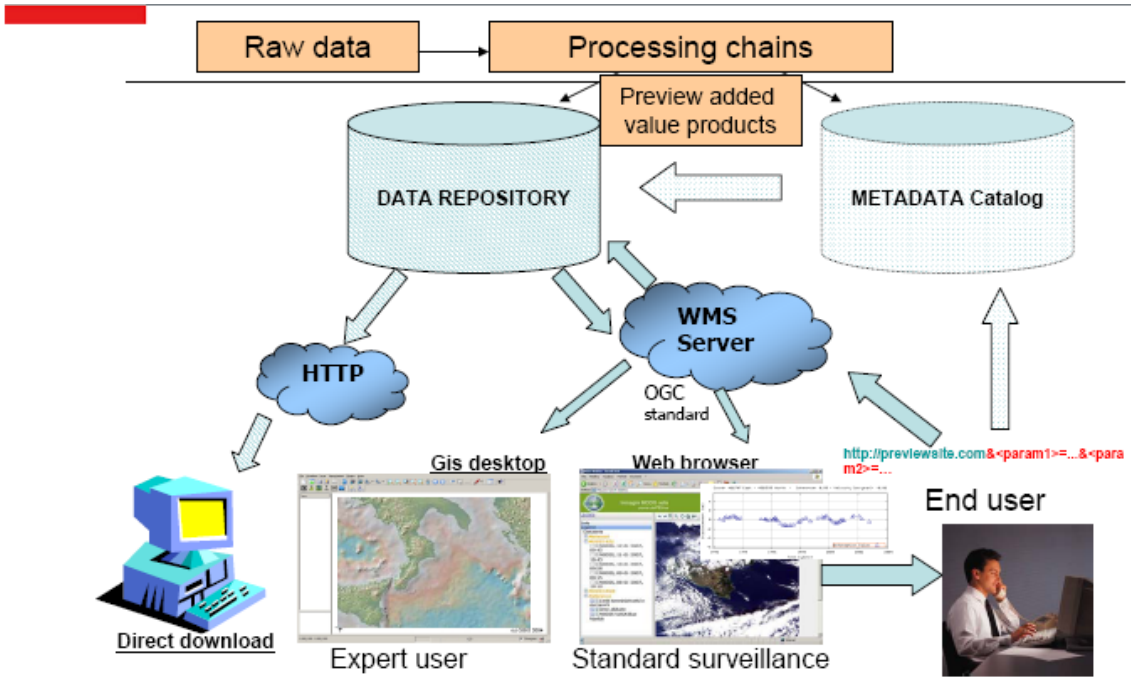


**Figure 23 - PREVIEW archive logical architecture**

### 4.34.3. Archive access and connection

External users can access the data (incl. raw data) using a web- portal. The data are protected by a login/password authentication mechanism.

### 4.34.4. Archive content

The PREVIEW project archive contains ~15 GB of EarthQuakes & Volcanoes platform data, Landslides platform data and Satellite from Europe in the time frame 1992 – 2005.

The project ended in December 2008.

### 4.34.5. Data curation and long-term preservation

PREVIEW applies backup and keeps to OpenGIS WMS/ISO:19128 2005 and OGC WMS. Data sets are stored in ESRI Shapefile and Geotiff for raster images.Metadata are compliant to ISO 19115:2003.

End of Project was in December 2008. So the long-term preservation is not clear.

## 4.35 RISK-EOS

### 4.35.1. Introduction

RISK-EOS is a network of European service providers delivering geo-information services to support the management of flood, fire and other risk throughout all phases: Prevention, Early warning, Crisis & Post crisis. The RISK-EOS services combine the use of satellite observation data with exogenous data and modelling techniques. The information in this section is collected from the questionnaire returned by Susana Martinez.

### 4.35.2. Archive systems

The archive data are satellite data (for provision of forest fire services in Spain), stored on a system of PC, external hard disks and DVD with backup copy and as a redundant and recovery system. The amount of data is 80 GB with expected increase of 50 GB per year.

### 4.35.3. Archive access and connection

There is no access for external users, as the repository not connected to the public network. Internally, the archive is accessed via password.

### 4.35.4. Archive content

The archive data are satellite data (Landsat, Spot, MSG, TERRA, Aqua). Formats of data are according to internal policies. Formats of meta-data are according to internal policy as well as INSPIRE compliant.

### 4.35.5. Data curation and long-term preservation

Upgrade of software and hardware, back-up and discarding of data is according to internal policy. No curation is in place to keep track of volatile information, no sustainability strategies are implemented. Sustainability of quality, and increase of quality, is reported, with main obstacles being storage capacity, security and standardization, the effort is one man month per year and 60 K€ per year.

## 4.36 SIB-ESS-C Data Repository

### 4.36.1. Introduction

The Department of Geoinformatics, Hydrology and Modelling (DGHM) is an innovative academic research institute at the Friedrich-Schiller University (FSU-Jena). Since it was established as part of the Institute of Geography (http://www.geogr.uni-jena.de) in 1994 it is carrying out research on how distributed geographical objects of a heterogeneous Natural Environment (NE-System) together with its related Human Dimension (HD-System) constitute respective NEHD-Systems can be classified, and how to analyse, model and regionalize their process driven interactive dynamics. Respective research projects carried out by the DGHM comprise (i) process analysis of complex geo-systems; (ii) management and modelling of spatial distributed heterogeneous geo-information; (iii) regional modelling of distributed water and solute transport; (iv) Integrated Water Resources Management Support Systems (IWRMSS) and (v) the regionalization concept of Response Units (RU). (see Website: http://www.geoinf.uni-jena.de).

Within this institute, the mission of the Siberian Earth System Science Cluster is to provide a web-based infrastructure and comprehensive information products derived from Earth Observation that support environmental and earth system research in Siberia. (see Website: http://www.sibessc.uni-jena.de/).

In the initial phase of the SIB-ESS-C project data sets and value-added products created within the SIBERIA-II project will be disseminated. A major goal of SIB-ESS-C is to continue product generation in order to build up time series for environmental monitoring and as input parameters for earth science models. During the course of the project we intend to gradually extend the set of products through our own production facility and through collaboration (from http://www.sibessc.uni-jena.de/).

### 4.36.2. Archive systems

The archive system of the University of Jena is used.

It has backup copy and recovery system and is no redundant system, with media store copy on hard disk.

### 4.36.3. Archive access and connection

The archive can be accessed via web site portal and access via gateway for registered users.

No raw data are provided to external users.

External progress capacity is planned through OGC WPS.

### 4.36.4. Archive content

Products provided cover central Siberia and have been created by a consortium of research institutions that joined forces in the FP 5 EU project SIBERIA-II (http://www.siberia2.uni-jena.de/) (Multi-Sensor Concepts for Greenhouse Gas Accounting of Northern Eurasia, EVG2-2001-00008).

The study region comprises a number of ecosystems in northern Eurasia ranging from the tundra, the boreal and temperate forests, mountainous areas and grasslands. The region is believed to play a critical role in global climate change and has been also defined as one of IGBP's http://www.igbp.kva.se/ Boreal transects representing a strong climate change hot spot in Northern Eurasia (from http://www.sibessc.uni-jena.de/sites/overview.html).

The actual amount of data is 50 GB, expected are 10 GB per year.

### 4.36.5. Data curation and long-term preservation

The data formats are OpenGIS GML, ESRI Shapefile EO format (CEOS, DIMAP...) and the associated metadata comply with ISO 19115&19139 standards.

For data discovery, the services are based on OGC (CSW, WMS) standards while OpenGIS (SLD, WMS) standards are used for content rendering.

Except for the use of these standards no particular data curation and long-term preservation policy has been mentioned.

## 4.37 Starlab Data Repository

### 4.37.1. Introduction

Starlab Space develops new and alternative methods for Remote Sensing applications. On the technology side, it develops new space sensors and mission concepts for environmental monitoring, including remote sensing applications of GPS and Galileo, as well as new radar concepts.

Starlab also implements and manages operational information services using space data for oil spill monitoring and water quality.

The information in this section is compiled from http://starlab.es/space and the questionnaire returned by Araceli Pi Figueroa (Starlab).

### 4.37.2. Archive systems

The archive system consists of a file server machine and MySQL and PostgreSQL databases. Data sets are stored on tape.

### 4.37.3. Archive access and connection

With appropriate tools users can access the data sets after authentication with password. The archive has Highspeed connectivity (GEANT, HiSEEN).

### 4.37.4. Archive content

The archive stores data sets from ENVISAT ASAR, MERIS and Ground Based Sensors of the last 2 years. Main application field is water. The yearly data growth is in the range of 2 GB.

### 4.37.5. Data curation and long-term preservation

The Starlab archive manages the data using backup, redundant systems and recovery systems.

## 4.38 Swisstopo Data Repository

### 4.38.1. Introduction

The GIS-infrastructure of Swisstopo is considered to produce and provide and represent geospatial data of Switzerland. Several data repositories for the production of geospatial data are in use or under construction. These are for example the GTDB for VECTOR25, TOPGIS for the TLM (Topographic Landscale Model), VEC (VECTOR200), Swissimage (DOP of Switzerland), GENIUS-DB (digital cartographic model). To provide the geospatial data, Swisstopo has a product repository called PRODAS (product database) and several Web-GIS-repositories. The GIS for the production are redundant and consist of three parts: the test-system, the integration-system and the production-system. The information in this section is collected from the questionnaire returned by Buogo Alain.

### 4.38.2. Archive systems

The Swisstopo data repository is located at the Computing Centre in Wabern and the Computing Centre BIT in Bern on HP servers running MS Windows 2003 Server / Linux SLES10 and with the following software packages: ESRI ArcSDE, ESRI ArcGIS Server, ESRI ImageServer and Oracle 10.2. The archive system has a backup copy, is a redundant and recovery system. Media store copy is on Disc and Tapes.

### 4.38.3. Archive access and connection

The archive can be accessed via web site portal using appropriate tools and access via gateway for selected users with passwords. Standard IT security mechanism has been employed to protect against unauthorised users. No raw data are provided to external users. There is a high speed network SWITCH.

### 4.38.4. Archive content

The data (Geology, Land Surface, Human Powered Mobility) is from Aerial, Ground Based Sensors and Reference Partners and has been collected since 1839. It covers National area. The actual amount of data is as follows: Vector Data: 1TB, Raster Data: 5 TB (not raw data), and 100TB Air Borne Images and is expected to grow by 30-50 TB per year.

### 4.38.5. Data curation and long-term preservation

The data formats are KML, INTERLIS, DXF, E00 and TIFF. ESRI Shapefile, OpenGIS GML and Web Map Context are under development. The associated metadata complies with ISO 19115:2003, ISO 19139, and INSPIRE compliant ISO 19115-Swiss Profile GM03 standards.

Data discovery services will be based on internal policies OGC (WMS, RSS, CSW) while internal policies OpenGIS WMS/ISO: 19128 2005 are used for content rendering. The services will be described using a WSDL document based on internal policies.

The long term access will be provided by migration of data and this is also work in progress.

## 4.39 TNO Data Repository

### 4.39.1. Introduction

DINOShop (www.dinoloket.nl) is the central portal to Data and Information of the Subsurface of the Netherlands. The DINO system is the central storage site for geoscientific data on the shallow and deep Dutch subsurface. The archive contains shallow and deep borings, groundwater data, cone penetration tests, geo-electrical measurements, results from geological, geochemical and geomechanical sample analysis, borehole measurements and seismic data.

### 4.39.2. Archive systems

The data repository is located in Utrecht on hardware subcontracted to external company running Windows and Linux, DBMS Oracle, services and applications in Java. The archive system has a backup copy, is a redundant and recovery system. Media store copy is on Disc and Tapes.

### 4.39.3. Archive access and connection

The archive can be accessed via web site portal and access via gateway for registered users.

There is a separate site for confidential data; public data available on subscription protected by email address and password. Raw data is provided to external users for download. There is a high speed network SURFnet.

### 4.39.4. Archive content

The Geological data is from Ground Based Sensors and has been collected since 1850. It covers National area. The actual amount of data is 20 GB and is expected to grow by 1 GB per year.

### 4.39.5. Data curation and long-term preservation

The data formats are ESRI Shapefile and OpenGIS GML. The associated metadata complies with ISO 19115-2, and ISO 19139 INSPIRE standards.

Data discovery services are based on OGC (WMS, CSW) while the services are described using INSPIRE. There is no long term access plan.

## 4.40 University of Helsinki Data Repository

### 4.40.1. Introduction

The Division of Atmospheric Sciences (http://www.atm.helsinki.fi/indexeng.html) of the University is involved in the following projects: CarboEurope, NitroEurope, IMECC and ICOS.

### 4.40.2. Archive systems

Biosphere-atmosphere interactions and atmospheric physics and chemistry data are stored in SMEAR station databases. The archive system has a backup copy, is a redundant and recovery system. There is no media store copy.

### 4.40.3. Archive access and connection

The archive can be accessed via web site portal freely but not via Gateway. University firewall protects against unauthorized access. Raw data is provided to external users on physical media. There is no high speed network.

### 4.40.4. Archive content

The Atmosphere/Weather data is from Ground Based Sensors and covers local area (less than National area) since 1955. The actual amount of data is unknown but is increasing each year.

### 4.40.5. Data curation and long-term preservation

The data formats along with their associated metadata are from those agreed in EU projects.

Data discovery services and services are also described based on internal policies while there are no content rendering standards. There is no long term access plan.

## 4.41 University of Strasbourg Data Repository

### 4.41.1. Introduction

SERTIT (Service Régional de Traitement d'Image et de Télédétection) is a valorisation service from the university developing and offering EO applications and services.

SERTIT, a remote sensing and image processing service was created in 1987 by Paul DE FRAIPONT within the ENSPS, Strasbourg, a graduate Engineering School. SERTIT's vocation is to format data and extract information from data produced by Earth Observation systems. As well as optical data it processes radar data from operational satellites: SPOT, IKONOS, LANDSAT, ERS 1-2, RADARSAT, ENVISAT.

SERTIT is also specialized in crisis remote sensing applications. It is supported by the European Space Agency (ESA) and has just been contracted by the French Space Agency (CNES) to produce added value EO derived products, as of January 2004 and for 3 years, within the framework of the International Charter 'Space and Major Catastrophes'.

SERTIT develops rapid mapping services (ESA GSE - Risk-eos, Respond - ESA EOMD).

The information in this section is compiled from the SERTIT website (http://sertit.u-strasbg.fr).

### 4.41.2. Archive systems

The archive system has a backup copy, is a redundant and recovery system. Media store copy is on LTO and HDD.

### 4.41.3. Archive access and connection

FTP access is granted, with some data access free and some data for registered users only. No raw data are given to external users.

No access via gateway. There is a high speed network RENATER.

### 4.41.4. Archive content

The data formats are OpenGIS GML & Web Map Context, and ESRI Shapefile EO format (CEOS, DIMAP...) and the associated metadata follow an internal policy compliant with INSPIRE recommendation.

For data discovery, the services are based on OGC (CSW, WMS) and OpenGIS CAT standards while OpenGIS (SLD, WMS) standards are used for content rendering.

Except for the use of these standards no particular data curation and long-term preservation policy has been mentioned.

### 4.41.5. Data curation and long-term preservation

They have not yet planned a specific data curation and long-term preservation policy, but however they do track volatile information.

## 4.42 WDC-Climate

### 4.42.1. Introduction

The mission of the WDCC corresponds with the basic tasks of Model and Data. Data for climate research are collected, stored and disseminated in order to serve the scientific community. M&D has experience in running international data centres. The WDCC is restricted to climate data products. No raw data storage for example from satellites or climate models is planned in the WDCC itself because raw data storage on a global basis is beyond the scope of the available facilities. Emphasis is spent on climate modelling and related data products. A close cooperation with thematically corresponding data centres like Earth observation, meteorology, oceanography, pale climate and environment is planned in order to establish a complete network for climate data. The WDCC is maintained by Model and Data, which is hosted at the Max-Planck-Institute for Meteorology, in cooperation with the German Climate Computing Centre (DKRZ). The information in this section is compiled from WDC-Climate website (www.wdc-climate.de) and the questionnaire returned by Michael Lautenschlager (WDC-C).

### 4.42.2. Archive systems

The WDC-C archive is based on Linux- and Sun-servers, Oracle RDBMS and Unitree HSM and considered OAIS recommendations. Selected data sets are stored on Magnetic tapes for long term.

The used meta data system is ISO 19115-2 ISO 19139 INSPIRE compliant.

### 4.42.3. Archive access and connection

External users can access WDC-C data sets (incl. selected raw data) via the web- portal: http://cera.wdc-climate.de. User access is password protected. Identification by certificates is planned for the future.

WDC-C has a high- speed internet connectivity using the DFN scientific network–Germany's National Research and Education Network.

Processing options for data format conversion and data volume reduction are offered.

### 4.42.4. Archive content

The Archive contains data sets of Atmosphere/Weather, Land Surface, Ecosystems, Water, Oceans, aerosols, chemistry, bio-geochemical cycles, and glaciers from a multitude of sources like Satellites, aerial, GBS, Statistic, numerical models. The focus is Europe but also world wide.

The data sets cover a time span of approximately 3000 years in time with emphasis on period 1860 – 2100. Actually the data sets comprise 360 TB. The data growth is expected at 1 PB/year.

### 4.42.5. Data curation and long-term preservation

WDC-C spends much effort for Data curation and long-term preservation. The sustainability strategy is described in http://www.pv2007.dlr.de/ . It operates backup and recovery systems. Data are presently stored in duplicate versions and regularly migrated to new storage media.

Beside bit-streams, format access libraries and metadata integrity WDC-C preserves numerical models together with execution scripts are preservation of but this might be not sufficient for really reprocess the data.

Upgrade of H/W and S/W must be downwards compatible if migration is not possible.

These efforts are described more in detail in http://ceramodel.wdc-climate.de/.

## 4.43 WDC-MARE and PANGAEA

### 4.43.1. Introduction

The WDC-MARE (World Data Centre for Marine Environmental Sciences) is aimed at collecting, scrutinizing, and disseminating data related to Global Change and earth system research in the fields of environmental oceanography, marine geosciences, and marine biology (see Website: http://www.wdc-mare.org).

It focuses on geo-referenced data using the PANGAEA (Publishing Network for Geoscientific & Environmental Data) information system as its long-term archive and publication unit (see WebSite: http://www.pangaea.de).

The WDC-MARE provides data management and services for several projects (14 completed, 11 in progress) like EUR-OCEANS, CARBOOCEAN, HERMES, etc.

### 4.43.2. Archive systems

The WDC-MARE is maintained by the Alfred Wegener Institute for Polar and Marine Research (AWI) in Bremerhaven (Germany), a research center of the Helmholtz Association, and the Centre for Marine Environmental Sciences (MARUM), University of Bremen (Germany).

The information system PANGAEA is used by WDC-MARE for archiving.

Two StorageTek Libraries SL8500 (each with 18 tape drives (LTO/3) and 3000 tapes (50-400 GB)) are used by WDC-MARE/PANGAEA for backup and archiving of large datasets and binary objects, e.g. from seismic, bathymetry, image material or modelling. The total capacity is up to 1200 TB mirrored with full redundancy in two buildings at different locations.

The PANGAEA technology is based on a three tiered client/server architecture with a data set cache:

- A relational database is the *back-end* and central archiving system using Sybase as the RDB management software on a multiprocessor computer.

- As *middleware*, an application server with open server components for import, retrieval and editing is operated. Components are encapsulated and use standard interfaces for communication.
- On the *front-end side* different clients provide access to the system. The graphical user interface (GUI) for data upload and metadata definitions is a proprietary application written in 4th Dimension (ACI). Middleware and front-end components follow a generic model to ensure a flexible functionality and easy modifications.

### 4.43.3. Archive access and connection

The strictly normalized data model follows the steps in science for gathering analytical data. Up to 12 tables can be used to define metadata which are related to the data during import. Data sets are stored in a pre-configuration *as defined by the author* in the database cache in ISO/XML format to assure fast access. Each data value can be geo-referenced in space and/or time, allowing individual extractions from the inventory. Data sets of up to some million items are stored in two tables (numeric, text) organized through an index tree. Larger data sets or binary objects are stored as files, linked to its meta description and geo-reference in Pangaea. As file and backup archive two tape drive systems (SL8500) are used in combination with a hard disc array. The system has an Internet connection of 1 Gbit and is connected with the HiSEEN/GÉANT network.

*Web services* and *clients* are provided for data retrieval, download and harvesting:

- PANGAEA Search engine with a map to set geographical constraints and visualization of the results;
- Data mining tool ART;
- Map server for profiling data PIM;
- Link definitions starting a query to provide dynamic lists of a specific part of the content;
- Portals to harvest the PANGAEA content with Open Archives Initiative (OAI) protocol using metadata in DIF or ISO-19139 format.

A Wiki is operated for the documentation of any technology, functionality and use around PANGAEA.

### 4.43.4. Archive content

The WDC is a multidisciplinary archive of Marine Environmental Science data from aerial, ground and water based operations. It covers all Europe, from geological times up to now. The amount of data is about 5 TB with an expected increase of 0,5 TB/year.

### 4.43.5. Data curation and long-term preservation

PANGAEA guarantees long-term availability of its content through a commitment of the operating institutions. Each dataset can be identified, shared, published and cited by using a Digital Object Identifier (DOI). In addition a citation for data collections can be defined on request. Citations are available through the catalogue of the German National Library of Science and Technology (TIBORDER). Archiving follows the Recommendations of the Commission on Professional Self Regulation in Science for safeguarding good scientific practice.

The system is operated in the sense of the *Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities* (http://oa.mpg.de/openaccess-berlin/berlindeclaration.html) which is a follow-up to the *Budapest Open Access Initiative* (http://www.soros.org/openaccess/).

The policy of data management and archiving follows the Principles and Responsibilities of ICSU World Data Centres (http://www.ngdc.noaa.gov/wdc/guide/gdsystema.html) and the OECD Principles and Guidelines for Access to Research Data from Public Funding (http://www.oecd.org/document/55/0,3343,en_2649_34293_38500791_1_1_1_37417,00.html).

## 4.44 ZALF Data Repository

### 4.44.1. Introduction

ZALF's (Centre of agricultural research, Muncheberg, Germany) mission is to do scientific research on ecosystems in agricultural landscapes and to develop ecologically and economically tenable land use systems. The division Landscape Information Systems (LIS) is a central facility coordinating and implementing knowledge and information management at ZALF. Core parts of this infrastructure are central data-Repositories, Web- services, Workflows und web-based research platforms. The information in this section is compiled from http://www.zalf.de/home_zalf/gb_index.html and the questionnaire returned by Uwe Heinrich (ZALF).

### 4.44.2. Archive systems

The archive is based on following technologies: Sun, Solaris, ESRI, files and Intel, MS 2003, ESRI SDE, SQL Server. data sets are long term stored on tape.

### 4.44.3. Archive access and connection

External users can access and download data sets (incl. raw data) via a web site portal. This interface is protected against unauthorized access by authentication (UID, password).

### 4.44.4. Archive content

The archive contains a multitude of regional data sets describing landscape: atmosphere/ weather, geology, land surface, social and human topics, cadastral, ecosystems, and water.

The data cover a time frame of partly 100 years but mainly of the last 10 years.

The actual data sets amount to 300 GB with an estimated growth of 100 GB per year.

### 4.44.5. Data curation and long-term preservation

ZALF operates a backup system. Important data sets are mirrored. Backup and migration are standard procedures. Meta data system is INSPIRE compliant. Data stored as ESRI ArcSDE, Cover, and Grid.

## 4.45 Summary of the surveyed digital repositories

The present section presents a global assessment of the main characteristics of the surveyed digital repositories, so as to verify their adequacy with the GENESI-fication requirements and technological choices made by GENESI-DR.

### 4.45.1. Issue « INSPIRE-compliant metadata »

ISO 19115:2003 and ISO 19139 standards are mainly used. Some repositories follow internal policy but they aim to be INSPIRE compliant. One uses FGDC. The contents exist consequently for a GENESI-fication of the metadata (RDF/XMLfiles) the step may be easy.

Indeed GENESI-DR information objects follow the ISO 19115 standard for describing geographic information and services and the INSPIRE implementing rules ISO 15836.

### 4.45.2. Issue « OAIS standard as a reference »

There is a majority of « No » followed by « partial compliancy ». Only two repositories consider that they follow the standard.

It underlines the fact that each repository has its own methodology of development and operation of the archive, although international efforts are made to help them.

GENESI-DR recommends the use of OAIS Information Model so as to use data preservation metadata associated to data products.

### 4.45.3. Issue « Access Policy »

Only four repositories answer « free access »; most of the answers are « registered users » and « selected users with password ».

The ways of delivering data products are by download and media, after registration and authentication.

Hence the necessity of authentified access; GENESI-DR fulfils this need by implementing the notion of virtual organizations enabling the definition of different access policies.

### 4.45.4. Issue « Access for external users »

Most of the archives have their own web portal. They rarely offer external process capacity on the data.

The need of processing capacity has been underlined by scientists and in this area GENESI-DR may bring an advantageous solution, with its Grid processing capabilities.

### 4.45.5. Issue «Physical archive management »

Back-up, recovery and redundancy have mostly become usual ways of proceeding for maintaining a certain level of sustainability.

### 4.45.6. Issues of archive sustainability « Maintain or increase the quality level »

One can say they are all aware of the necessity and will do their best. But they have not detailed their strategy. It shows that there is a lack of on-the-shelf methodologies or tools (DNA4.4 and JRA2 results and other EU projects as CASPAR, PARSE.Insight, …) could be useful but easily implementable solutions are still necessary. But they all mention lack of time, funding and people, so the way of financing sustainability is still the key issue. They are lucky when taking part in an international program which can bring some funding.

### 4.45.7. Issue « Any standard for service description and discovery »

Although several answers do not enable to know the tendency, as the answer is « none » or « internal policy », the remaining answers underline the use of OpenGIS OGC or WSDL for service description and .OGC CSW or WMS for service discovery.

GENESI-DR enables to describe a service and provides in particular a metadata called « dclite4g:onlineResource » which is a reference to an online resource which provides direct access to the service. GENESI-DR vocabulary defines identifier URLs for different commonly-used services - WMS, WCS, LDAP, FTP among others. This capability enables to describe services which are available via GENESI-DR. Consequently, the archives using OGC standards could integrate their free accessible application in GENESI-DR.

### 4.45.8. Issue « Any standard for data format »

The answers underline the variety of data and metadata formats. Examples of formats are: OpengGIS Web Map Context ESRI Shapefile, Compact RINEX format, OpenGIS GML ESRI Shapefile, ESRI ArcSDE, Cover, Grid, OpenGIS GML, EO format (CEOS, DIMAP...), data accessible as SensorML, WMO (BUFR, GRIB) Standards, HDF Standards, ESRI ArcSDE, Cover, Grid.

However data formats are transparent since GENESI-DR provides access to the data products but does not provide visualization tools (out of the scope of the project). It is assumed that end-users use their own visualization facilities for displaying data.

### 4.45.9. Issue « Connection to a high-speed network »

Half of the archives are connected to a high-speed network (HiSEEN/GÉANT, NREN, EUMETCast/GEONATCast, GTS/WIS); the other half is stand-alone.

# 5. CONCLUSIONS

## 5.1 Infrastructures

The existence of high-speed networks such as GEANT or HIGHSEEN and the implementation of grid-based infrastructures such as G-POD provide the end-users with the possibility of quick access and bulk processing using facilities which do not have to be necessarily at their own site.

The development of the NREN also enables to have important network coverage for the exchange of data in the educational world; consequently, one can envisage having the possibility of defining one core node integrated within GENESI-DR for spreading data results according to MoM if a NREN is willing to integrate the federation.

The ES requirements analysis and projects survey described in the present document (see 3) shows how e-infrastructure, data centric and interoperability-based projects like GENESI-DR represent a most comprehensive solution to the ES needs.

Interoperability, data heterogeneity management and multi-disciplinarity requirements are in fact met and considered in the appropriate direction.

However, there is still room for improvement in order to face and satisfy the dynamically emerging ES needs.

For example, a real information system based on the latest technology is becoming necessary to provide decision-makers at all levels (local to European) with real-time data, thus allowing them to make immediate and life-saving decisions especially in cases of emergencies (disaster management etc.).

The enhancement paths are identified such as maximization of the interoperability, operativity and inter-disciplinary collaboration.

## 5.2 Digital Repositories

Archives with few services (i.e. services dedicated to orders and deliveries) may be interested by being integrated into GENESI-DR, as GENESI-DR offers discovery and access facilities and the possibility of integrating applications which may process the offered data products and an ingest facility for the resulting data products.

For archives with adding value services (such as specific transformation processing) which may be interested by being integrated into GENESI-DR, they will have to provide efforts for making their services communicate with GENESI-DR services, i.e. by implementing the GENESI-DR Opensearch-based protocol.

## 5.3 General

This document - a survey of the existing and planned infrastructures on data archive and dissemination of Europe - proves the great treasure of Geo-data in Europe distributed over a multitude of repositories. The low participation in answering the questionnaire (2.4) shows that digital repository owners with only little involvement in answering the questions would also put little time in promoting the conversion of their archives to modern methods (reasons: time, money, motivation). So it is necessary to show repository owners the advantage of switching to modern methods of data archiving and curation and to dock to an appropriate infrastructure (e.g. GENESI-DR). Promoting organizations could support this motivation (4.45.6).

This existing potential for scientific use is only partially utilized:

- Multitude of data policies hampers the easy utilization (4.45.3);
- Data only accessible via specialized web interfaces (4.45.4);
- User confronted with a multitude of data formats (4.45.8);
- Only a half of the repositories is connected via highspeed networks (4.45.9).

At the same time there is the risk of loss of data sets:

- It is true that the long term capabilities are on a good way but,
- The majority of repositories is not OAIS compliant (4.45.2, DRs not compliant to OAIS will be vulnerable in the longer term (more expensive, proprietary SW developments instead of using common software for OAIS issues, etc.)) and
- Work on archive sustainability is still hesitating (4.45.6).

But the digital repositories are in the phase of change. GENESI-DR (or similar infrastructures) can contribute by providing:

- A relatively easy integration of metadata (4.45.1) and compliance to INSPIRE;
- Examples for the usage of OAIS standard;
- Easy access mechanisms presenting different data policies;
- Solutions for archive sustainability and use of established standards of service;
- Processing capacity;
- Archiving capacity.

# Annex A. References

[1] GENESI-DR-NA1-DEL- DNA1.10 - GENESI-DR Description of Work, v2.0 (28/05/2009)

[2] INSPIRE State of Play 2007: http://inspire.jrc.ec.europa.eu/state_of_play.cfm)

[3] ISO 14721:2003, “Space Data and Information Transfer Systems - Open Archival Information System - Reference Model”, Edition 1, February 2003

[4] CCSDS 650.0-B-1., “Reference Model for an Open Archival Information System (OAIS)” – CCSDS Blue Book, Issue 1, January 2002

[5] ESA – How to get Earth Observation data : http://www.esa.int/esaEO/SEM12R1VQUD_index_0.html

[6] ESA Catalogues overview: http://earth.esa.int/resources/catalogues/

[7] ENVISAT Data Dissemination System: http://dwlinkdvb.esrin.esa.it/DDS/

[8] ESA and TPM Data Products overview: http://earth.esa.int/dataproducts/

[9] Standard Archiving Format for Europe: http://earth.esa.int/SAFE/

[10] European Strategy for Long term EO data preservation and access, ESA/PBEO/DOSTAG(2007)2, 8 October 2007.

[11] M. Albani, V. Beruti, M. Duplaa, C. Giguere, E. Mikusch, M. Serra, “Long Term Preservation of Earth Observation Space Data - European LTDP Common Guidelines” available at http://earth.esa.int/gscb/ltdp/EuropeanLTDPCommonGuidelines_DraftV2.pdf

[12] F. Coccoluto, C. Boer, A. Kaiser-Weiss, DNA4.2, Description of European Archiving Infrastructure, GENESI-DR-NA4-DEL-DNA4.2, 2008-11-03

[13] C. Boer, A. Kaiser-Weiss, K.-D. Missling, M. Serra, DNA4.4, European Concepts for Archive sustainability, GENESI-DR-NA4-DEL-DNA4.4, 2009-02-02

[14] F. Brito, F. Pacini, . Rossi, P. Goncalves, ESDR Minimum Requirements and GENESI-fication Guide, GENESI-DR-SA2-DEL-DSA2.4, 2008-09-02

[15] www.unece.org/env/pp/documents/cep43e.pdf

[16] www.ec-gis.org/inspire/

[17] A. Parrini, DSA2.2, Architectural Design, GENESI-DR-SA2-DEL-DSA2.2-v2.0, 09/07/2009

# Annex B. Figures and Tables

## B.1 List of Figures



## B.2 List of Tables

# Annex C. Terminology

The following terms and definitions apply.

## C.1 Definitions of Terms

| Term | Definition |
|---|---|
| **Access control system** | An access control system is a system which enables an authority to control access to areas and resources in a given physical facility or computer-based information system. An access control system, within the field of physical security, is generally seen as the second layer in the security of a physical structure |
| **Archive** | An archive is a collection of historical records, and the location in which the collection is kept. Archives contain records (primary source documents) which have accumulated over the course of an individual or organization's lifetime. |
| **Asset** | An asset is an object to which access is to be controlled. In the context of an Information Systemt this is most likely to be a data file, but could also be a web page or a Web Service. |
| **Authentication** | Authentication is the act of confirming the identity of a person. |
| **Author/distributor** | An author is defined both as the person or Institution who originates or gives existence to anything and that authorship determines responsibility for what is created. |
| **Authorisation** | Authorization is the function of specifying access rights to resources, which is related to information security and computer security in general and to access control in particular. More formally, "to authorize" is to define access policy. |
| **Backup** | In information technology, a backup or the process of backing up refer to making copies of data so that these additional copies may be used to restore the original after a data loss event. These additional copies are typically called "backups." The verb is back up in two words, whereas the noun is backup (often used like an adjective in compound nouns). Backups are useful primarily for two purposes. The first is to restore a state following a disaster (called disaster recovery). The second is to restore small numbers of files after they have been accidentally deleted or corrupted. |

| | |
|---|---|
| **Browser** | A web browser is a software application for retrieving, presenting, and traversing information resources on the World Wide Web. An information resource is identified by a Uniform Resource Identifier (URI) and may be a web page, image, video, or other piece of content.[1] Hyperlinks present in resources enable users to easily navigate their browsers to related resources.Although browsers are primarily intended to access the World Wide Web, they can also be used to access information provided by web servers in private networks or files in file systems.The major web browsers are Windows Internet Explorer, Mozilla Firefox, Apple Safari, Google Chrome, and Opera. |
| **Catalogues** | The catalog of a database instance consists of metadata in which definitions of database objects such as base tables, views (virtual tables), synonyms, value ranges, indexes, users, and user groups are stored |
| **Consumer** | Is the data re/user. |
| **Convention** | A convention is a set of agreed, stipulated or generally accepted standards, norms or criteria |
| **Curation** | Curation deals with the capacity of a system to offer and promote a reliable access to its data over time and assure their correct use. |
| **Custodian** | Is the data centre managing the data; |
| **Data** | Are observational data, model results, images. |
| **Data Acquisition** | In computer data processing, data acquisition is the sampling of real world physical conditions and conversion of the resulting samples into digital numeric values that can be manipulated by a computer. Sometimes abbreviated DAQ, data acquisition typically involves the conversion of analog signals and waveforms into digital values and processing the values to obtain desired information. The components of data acquisition systems include: Sensors that convert physical parameters to electrical signals. Signal conditioning circuitry to coerce sensor signals into a form that can be converted to digital values. Analog-to-digital converters, which convert conditioned sensor signals to digital values. |

| | |
|---|---|
| **Data Management** | Data Resource Management is the development and execution of architectures, policies, practices and procedures that properly manage the full data lifecycle needs of an enterprise |
| **Data Model** | Data models support data and computer systems by providing the definition and format of data. If this is done consistently across systems then compatibility of data can be achieved. The possibility of including the information in the data file is depending on data model adopted. |
| **Data Policy** | A policy is typically described as a deliberate plan of action to guide decisions and achieve rational outcome(s). In the project it is defined as the guide for accessing data. |
| **Data Processing** | Computer data processing is any process that uses a computer program to enter data and summarise, analyse or otherwise convert data into usable information. The process may be automated and run on a computer. It involves recording, analysing, sorting, summarising, calculating, disseminating and storing data. |
| **Digital Library** | A digital library is a library in which collections are stored in digital formats (as opposed to print, microform, or other media) and accessible by computers. The digital content may be stored locally, or accessed remotely via computer networks. A digital library is a type of information retrieval system |
| **Data Repository or Digital Repository** | A digital repository is an organization, which might be virtual, that comprehensively collects, manages and preserves for the long term rich digital content, and offers to its user communities specialized functionality on that content, of measurable quality and according to codified policies. An ES/EO repository contains Earth Science data. |
| **Discovery Service** | Service discovery protocols are network protocols which allow automatic detection of devices and services offered by these devices on a computer network. |
| **Distributed Data Base** | A distributed database or system is where data can be located at various geographically distributed nodes but still be accessible through one unique portal or hub. |

| | |
|---|---|
| **Download** | In networks, to download means to receive data to a local system from a remote system, or to initiate such a data transfer. Examples of a remote system might from which a download might be performed include a webserver, FTP server, email server, or other similar systems. A download can mean either any file that is offered for downloading or that has been downloaded, the process of receiving such a file |
| **Earth Science** | Earth science (also known as geoscience, the geosciences or the Earth sciences), is an all-embracing term for the sciences related to the planet Earth |
| **Format** | A format usually refers to the set-up, plan, or design of a thing. A file format is a particular way that information is encoded for storage in a computer file |
| **Geospatial** | Geospatial is a term widely used to describe the combination of spatial software and analytical methods with terrestrial or geographic datasets |
| **Image Data** | Are images providing information on various aspects of the environment |
| **Information** | Information as a concept has many meanings, from everyday usage to technical settings. The concept of information is closely related to notions of constraint, communication, control, data, form, instruction, knowledge, meaning, mental stimulus, pattern, perception, and representation |
| **Information System** | In a broad sense, the term Information Systems (IS) refers to the interaction between people, processes, and technology. This interaction can occur within or across organizational boundaries. An information system is not only the technology an organization uses, but also the way in which the organization's people interact with the technology and the way in which the technology works with the organization's business processes. Information systems are distinct from information technology in that an information system has an information technology component that interacts with the people and processes components. |

| | |
|---|---|
| **Infrastructure** | Infrastructure can be defined as the basic physical and organizational structures needed for the operation of a society or enterprise, or the services and facilities necessary for an economy to function |
| **Infrastructure for Spatial Information** | "infrastructure for spatial information" means metadata, spatial data sets and spatial data services; network services and technologies; agreements on sharing, access and use; and coordination and monitoring mechanisms, processes and procedures, established, operated or made available in accordance with this Directive; |
| **Ingestion** | Ingestion can also refer to a mechanism picking up something and making it enter an internal hollow of that mechanism |
| **Integration** | System integration, the process by which smaller pieces of software are brought together to form a larger piece of software that was designed to solve a problem Digital integration, in computer science, allows data from one device or software to be read or manipulated by another, resulting in ease of use; see also XML |
| **Interoperability** | Is the capability to communicate, execute programs, or transfer data among various functional units in a manner that requires the user to have little or no knowledge of the unique characteristics of those units. |
| **Interpretability** | interpretability is a relation between formal theories that expresses the possibility of interpreting or translating one into the other. |
| **Library** | A library is a collection of sources, resources, and services, and the structure in which it is housed |
| **License** | Is to give permission. A license is the document demonstrating that permission. A license may be granted by a party ("licensor") to another party ("licensee") as an element of an agreement between those parties. A shorthand definition of a license is "a promise (by the licensor) not to sue (the licensee) providing the terms of the license are respected". |

| | |
|---|---|
| **Mapping** | Mapping is the process of creating data element mappings between two distinct data models. Data mapping is used as a first step for a wide variety of data integration tasks including:<br>- Data transformation or data mediation between a data source and a destination<br>- Identification of data relationships as part of data lineage analysis<br>- Discovery of hidden sensitive data such as the last four digits social security number hidden in another user id as part of a data masking or de-identification project<br>- Consolidation of multiple databases into a single data base and identifying redundant columns of data for consolidation or elimination |
| **Metadata** | are data about data sets: information about observation, model or image data sets, the related attribute data such as type of data, location, date/time and unit of measurement, accuracy, precision or method of measurement, structure or ownership of the data. |
| **Metadata Model** | Meta-modeling is the analysis, construction and development of the frames, rules, constraints, models and theories applicable and useful for the modeling in a predefined class of problems. |
| **Migration** | Migration is the process of transferring data(softwares between storage types, formats, or computer systems. |
| **Model** | it corresponds to the representation or description designed to show the main object or workings of a concept. |
| **Model Data** | Are estimates of marine environmental variables from model outputs (including analyses). |
| **Naming Convention** | A naming convention is a set of rules for choosing the character sequence to be used for identifiers which denote variables, types and functions etc. in source code and documentation. |

| | |
|---|---|
| **Network** | A computer network or a group of interconnected computers, which enables computers to communicate with each other and share resources and information.<br>There are different types of networks from a small covered area to a world-wide covered area:<br>• Personal Area Network (PAN);<br>• Local Area Network (LAN);<br>• Campus Area Network (CAN);<br>• Metropolitan Area Network (MAN);<br>• Wide Area Network (WAN);<br>• Global Area Network (GAN);<br>• Virtual Private Network (VPN);<br>• Internetwork (Intranet, Extranet, Internet). |
| **Observation Data** | Are measurements or observations (in situ and remote sensing) of marine environmental variables. In order to interpret and process these data, related attribute data such as type of data, location, time and unit of measurement are also required. |
| **Originator** | Is the data producer |
| **Ownership** | Ownership is the state or fact of exclusive rights and control over property, including intellectual property. |
| **Password** | A password is a secret word or string of characters that is used for authentication, to prove identity or gain access to a resource |
| **Policy** | Is a set of rules that determine access to an asset. Policies may be thought of as locks that are put in place to protect assets. |
| **Preservation** | Long-term (long enough to be concerned with the impacts of changing technologies, including support for new media and data formats, or with a changing user community), error-free storage of digital information, with means for retrieval and interpretation, for the entire time span the information is required for. |
| **Product** | Are aggregated or combined sets of observation data, model data, images or meta-data (including web services) |
| **Protocol** | A protocol is a set of rules which is used by computers to communicate with each other across a network. A protocol is a convention or standard that controls or enables the connection, communication, and data transfer between computing endpoints. In its simplest form, a protocol can be defined as the rules governing the syntax, semantics, and synchronization of communication. |

| | |
|---|---|
| **Quality Flag** | Are alphanumeric indicators added to each numerical data at the end of the quality checks procedures. |
| **Recovery** | Process, policies and procedures developed to restore lost data. |
| **Redundancy** | Duplication of critical components of a system with the intention of increasing reliability of the system, usually in the case of a backup or fail-safe. |
| **Rendering** | To convert any coded content to the required format for display or printing. |
| **Repository (Digital)** | An organization, which might be virtual, that comprehensively collects, manages and preserves for the long term rich digital content, and offers to its user communities specialized functionality on that content, of measurable quality and according to codified policies. |
| **Reprocessing** | Duplication of critical components of a system with the intention of increasing reliability of the system, usually in the case of a backup or fail-safe. |
| **Right Management** | Digital rights management (DRM) is a generic term for access control technologies that can be used by hardware manufacturers, publishers, copyright holders and individuals to try to impose limitations on the usage of digital content and devices. |
| **Role** | Is a property of a user that determines the criteria that must be satisfied before access to an asset is granted. In many ways it can be thought of as a key that a user possesses |
| **Sensors** | A sensor is a device that measures a physical quantity and converts it into a signal which can be read by an observer or by an instrument. |
| **Server** | A server is an application running on a computer that delivers a service. |
| **Service** | The operations which may be performed, by invoking a computer application, on the spatial data contained in spatial data sets or on the related metadata |
| **Software** | software is a general term used to describe the role that computer programs, procedures and documentation play in a computer system |
| **Spatial Data** | Any data with a direct or indirect reference to a specific location or geographical area |

| | |
|---|---|
| **Spatial Data Set** | An identifiable collection of spatial data |
| **Upload** | uploading, is the sending of data from a local system to a remote system, such as a server, with the intent that the remote system should save a copy of whatever is being transfered, or the initiation of such a process |
| **User** | A user is a person who uses a computer or Internet service. A user may have a user account that identifies the user by a username |
| **Versioning** | Is the process of assigning either unique version names or unique version numbers to unique states of computer software. |
| **Virtual Organisation** | Is a group who shares the same computing resources |
| **Vocabularies** | The same variable is sometimes named in different ways according to the repositories or the applications. Common vocabularies are developed in many international initiatives, such as GEMET (promoted by INSPIRE as a multilingual thesaurus), UNIDATA, SeaDataNet, Marine Metadata Initiative (MMI). |

## C.2 Acronyms and Abbreviations

| | |
|---|---|
| ACOnet | Austrian Academic Computer NETwork |
| AMS | Archive Management System at ESA |
| ASCII Medatlas | a standard exchange format for the Mediterranean and Black Sea community |
| ASI | Italian Space Agency |
| BdMS | Bureau des Metadata et Services or Metadata and Services Register |
| BSPO | Belgian Science Policy Office |
| CD-R | Compact Disc – Read only |
| CEOS | Committee on Earth Observation Satellites |
| CESNET | Czech Educational and Scientific NETwork |
| CLS | Collecte Localisation Satellites |
| CNES | Centre National d'Etudes Spatiales (French Space Agency) |
| CNIG | Centro Nacional de Informacion Geografica |
| DANTE | Delivery of Advanced Network Technology to Europe |
| DDS | Data Dissemination System |
| DEM | Digital Elevation Model |
| DFN | Deutsches Forschungsnetz |

| | |
|---|---|
| DIF | Data Interchange Format |
| DIGEST | DIgital Geographic Information Exchange Standard |
| DIMAP | Format used for SPOT images |
| DIMS | Data Information and Management System |
| DLR | Deutschen Zentrum für Luft- und Raumfahrt (German Space Agency) |
| DOI | Digital Object Identifier |
| DRS | Data Request Server |
| DSM | Digital Surface Model |
| DTM | Digital Terrain Modeling |
| DVD-R | Digital Versatile Disc – Read only |
| DWDM | Dense Wavelength Division Multiplexing |
| EC | European Commission |
| EEA | European Environment Agency |
| EGEE | Enabling Grids for E-sciencE |
| ENEA | Ente per le Nuove Tecnologie, l'Energia e l'Ambiente (Italian National Agency for New technology, Energy and the Environment) |
| ENVISAT | ENVIronmental SATellite |
| EO | Earth Observation |
| ERS | European Remote Sensing |
| ES | Earth Science |
| ESA | European Space Agency |
| ESRIShapefile | Environmental Systems Research Institute shapefile |
| EUFAR | EUropean Fleet for Airborne Research |
| EUMETSAT | European Organisation for the Exploitation of Meteorological Satellites |
| EUSC | European Union Satellite Centre |
| FAT32 | File Allocation Table - 32 |
| FGS | Facility Ground Segment |
| FP | Framework Programme |
| FTP | File Transfer Protocol |
| GARR | Gestione Ampliamento Rete Ricerca |
| GB | Giga Bytes |
| GÉANT | pan-European research and education network |
| GENESI-DR | Ground European Network for Earth Science Interoperations – Digital Repositories |
| GEOSS | Global Earth Observation System of Systems |

| | |
|---|---|
| GIS | Geographical Information System |
| GMES | Global Monitoring for Environment and Security |
| GRNET | Greek Research and Technology Network |
| GSC | GMES Space Component |
| G-POD | Grid-Processing On Demand |
| GPS | Global Positioning System |
| GTS/WIS | Global Telecommunication System/World Meteorological Organisation Information System |
| GUI | Graphical User Interface |
| HiSEEN | High-Speed ESA Earth observation Network |
| HMA | Heterogeneous Mission Accessibility |
| HTTP | HyperText Transfer Protocol |
| ICCM | Instituto Canario de Ciencias Marinas |
| IDN | International Directory Network (NASA for CEOS) |
| IFREMER | Institut Français de Recherche pour l'Exploitation de la Mer (French Research Institute for Sea Exploitation) |
| INSAR | SAR Interferometry |
| INSPIRE | INfrastructure for SPatial InfoRmation in Europe |
| INTEGRAL | Interferometric Evaluation of Glacier Rheology and Alterations |
| IOC/IODE | Intergovernmental Oceanographic Commission/International Oceanographic Data and Information Exchange |
| IP | Internet Protocol |
| IRS | Indian Remote Sensing Satellite |
| ISO | International Organization for Standardization |
| JRC - CID | Joint Research Center - Community Image Data |
| KML | Keyhole Mark-up Language |
| KSAT | Kongsberg Satellite Services |
| LTDP | Long Term Data Preservation |
| LTO-4 | Linear Tape Open |
| Mbone | Multicast backbone |
| MMFE | Multi Mission Facility infrastructure Elements |
| MMFI | Multi-Mission Facility Infrastructure at ESA |
| MMOHS | Multi Mission Order Handling System |
| MRI | Ministry for Resources and Infrastructure - Malta |
| MSG | Meteosat Second Generation |
| MTG | Meteosat Third Generation |
| MUIS | Multi-mission User Interface System at ESA |

| | |
|---|---|
| NA4 | Networking Activity 4 |
| NASA | National Aeronautics and Space Administration |
| NERSC | Nansen Environmental and Remote Sensing Center |
| NetCDF | Network Common Data Format |
| NILU | Norsk Institutt for Luftforskning (Norvegian Institute for Aerospace Research) |
| NINA | Norwegian Institute for NAture research |
| NLS | National Land Survey of Finland |
| NOA | National Observatory of Athens |
| NODC | National Oceanographic Data Centres |
| NREN | National Research and Education Network |
| OAI | Open Archives Initiative |
| OAI | Open Archives Initiative |
| OAIS | Open Archival Information System |
| OASIS | Optimising Access to SPOT Infrastructure for Science |
| ODV | Ocean Data View |
| OGC | Open Geospatial Consortium |
| OpenGIS | trademark of OGC (SLD=Style Layer Descriptor ; WMS=Web Map Service ; WMC=Web Map Context ;WSC=Web Service Common ; WFS=Web Feature Service WCS=Web Coverage Service ; ORM=OpenGIS Reference Model;CAT=CATalogue ; CSW=Catalogue Service for Web; GML= Geography Markup Language) |
| PAC | Processing and Archiving Centre – from ESA |
| PAF | Processing and Archiving Facility – from ESA |
| PANGAEA | Publishing Network for Geoscientific and Environmental Data |
| PB | PentaBytes |
| PC | Personal Computer |
| PFD | Product Formatting and Delivery |
| PIONIER | Polish Optical Internet – Advanced Applications, Services and Technologies for Information Society |
| POH | Product Ordering System |
| PSM | Processing System Management |
| RAID | Redundant Array of Independent Disks |
| RDB | Remote Data Base |
| RENATER | Réseau national de télécommunications pour la technologie, l'enseignement et la recherche |
| RSS | Really Simple Syndication |
| SAFE | Standard Archiving Format for Europe |

| | |
|---|---|
| SAN | Storage Area Network |
| SANET | Slovak Academic NETwork |
| SAR | Synthetic Aperture Radar |
| SDTS | Spatial Data Transfer Standard |
| SeaDataNet | a Pan-European Infrastructure for Ocean and Marine Data Management |
| SERAD | Services de Référencement et d'Archivage de Données or Data Recording and Archiving Service |
| SIPAD | Système d'Information, de Préservation et d'Accès aux Données (Information System for Data Preservation and Access) |
| SNSB | Swedish National Space Board |
| SPOT | Satellite Pour l'Observation de la Terre (Satellite for Earth Observation) |
| SRTM | Shuttle Radar Topography Mission |
| SST | Sea Surface Temperatures |
| STAF | Service de Transfert et d'Archivage de Fichier (File Transfer and Storing Service) |
| TA | Transnational Access |
| TB | Tera Bytes |
| TERENA | Trans-European Research and Education Networking Association |
| TIFF | Tagged Image File Format |
| UDDI | Universal Description Discovery and Integration |
| USB | Universal Serial Bus |
| VPN | Virtual Private Network |
| WDC-MARE | World Data Center for Marine Environmental Sciences |
| WIS | WMO Information System |
| WMO | World Meteorological Organization |
| WMO BUFR | Binary Universal Format for Representation for distributing SeaWinds data products |
| WMO GRIB | General Regularly-distributed Information |
| WSIL | Web Services Inspection Language |
| XML | eXtensible Markup Language |

# Annex D. European Infrastructure Questionnaire

## D.1 Questionnaire

*GENESI-DR is conducting a survey concerning the existing archiving infrastructures used for geo-spatial data around Europe.*
*Please help us by answering some questions concerning your archiving infrastructure(s).*

*The survey includes the following data:*
*1. Contact information,*
*2. General geo-spatial infrastructure information (projects),*
*3. Geo-spatial archive information (repositories, standards, curation, etc.).*

*The following questions are part of questionnaire.*
*Fields marked with * are mandatory.*

***Contact information***

- Name :_____
- Phone Number :_____
- E-mail address* :_____
- Organization Name :_____
- Organization Type
  - Earth Observation
  - Mapping agency
  - Research Institute
  - University
  - National agency (space, marine, forestry, statistics…)
  - Image Processing
  - Other, please specify below :_____
- Country (main site)* :_____
- Area of Interest
  - World
  - Europe
  - Part of Europe
  - Nation
  - Local
- Web site :_____

***Projects 1/2***

- Are you involved in a project using geo-spatial data ?
  - Yes
  - No

***Projects 2/2***

*Please describe the projects you are involved using geo-spatial data.*

- Project Name* :_____
- Project web site :_____
- Description :_____
  _____
  _____
- Your role* : _____
- Data Providers:

| Name | Contact |
|---|---|
| | |
| | |
| | |

***Repositories 1/2***

- Are there geo-spatial data repositories in your organization?
  - Yes
  - No

***Repositories 2/2***

*Please describe your data repositories.*
*Note: In the following questions we will assume that each repository has homogeneous characteristics in contents, access policy, etc. If this is not the case, please spread the description in several records or refer to the main amount of data.*

**Repository Description**

- Repository Location :_____
- Repository Identifier (mission, sensors) :_____

- Describe your main repository platform (hardware, software, database)
  :_____

- Do you have backup copy (near-line, off-line) of your data repository?
  - Yes
  - No

- Do you have a redundant system?
  - Yes
  - No

- Do you have a recovery system?
  - Yes
  - No

- Which media are you using to store off-line copy of your data :_____

**Repository Contents**

- What it the subject of your data repository
  - Atmosphere/Weather
  - Geology
  - Land Surface
  - Social and Human Topics

- ❑ Cadastral
- ❑ Ecosystems
- ❑ Energy
- ❑ Water
- ❑ Other, please specify below :_____

- What it the source of your data
  - ❑ Satellite
  - ❑ Aerial
  - ❑ Ground Based Sensors
  - ❑ Statistics
  - ❑ Other, please specify below :_____

- What is the covered area ?
  - ❑ Europe
  - ❑ Regional (international area)
  - ❑ National
  - ❑ Local (less than national area)

- What is the covered time ? : _____

- What is the actual amount of data (GB) in your repository? : _____

- What is the amount of data (GB/year) you expect for the future? :_____

**Repository Access**

- Do you provide access to your repository to external users?
  - ❑ No
  - ❑ Yes, with appropriate tools
  - ❑ Yes, web site portal
  - ❑ Other/Planned, please specify below :_____

- Do you provide any original raw data to external users?
  - ❑ No
  - ❑ Yes, via webservice
  - ❑ Yes, download
  - ❑ Yes, physical media

- How do you protect your repository from unauthorized users?
  :_____
- Is your repository accessible via a gateway?
  - ❑ Yes
  - ❑ No

- Is your repository connected to a high-speed network?
  - ❑ No
  - ❑ Yes HiSEEN/GÉANT
  - ❑ Yes, GTS/WIS
  - ❑ Yes, EUMETCast/GEONATCast
  - ❑ Other/Planned, please specify below :_____

- Do you provide external process capacity for your archive?
  - ❑ None

- ❑ Yes/Planned, please specify below :_____

➢ What is the applied access policy?
- ❑ Free
- ❑ Registered user
- ❑ Selected users with password
- ❑ Other/Planned, please specify below :_____

***Standards***

➢ Is your archive compliant with the OAIS reference model?
- ❑ Yes
- ❑ No
- ❑ Partially

➢ Do you follow any standard for content rendering?
- ❑ None
- ❑ Internal policy, please describe it
- ❑ OpenGIS SLD
- ❑ OpenGIS WMS/ISO:19128 2005
- ❑ ISO 19117:2005
- ❑ Other standard, please specify below :_____

➢ Do you follow any standard for service discovery?
- ❑ None
- ❑ Internal policy, please describe it
- ❑ OGC CSW
- ❑ OGC WMS
- ❑ OpenGIS CAT
- ❑ UDDI
- ❑ RSS
- ❑ WSIL
- ❑ Other standard, please specify below :_____

➢ Do you follow any standard for service description?
- ❑ None
- ❑ Internal policy, please describe it
- ❑ OpenGIS ORM
- ❑ OpenGIS OGC Common
- ❑ Other standard, please specify below :_____

➢ Do you follow any standard for data exchange?
- ❑ None
- ❑ Internal policy, please describe it
- ❑ DIGEST
- ❑ SDTS
- ❑ OGC CSW
- ❑ OpenGIS WFS/ISO 19142
- ❑ OpenGIS WCS
- ❑ OpenGIS Filter/ISO 19143
- ❑ OpenGIS CT
- ❑ Other standard, please specify below :_____

➢ Do you follow any standard for the data format?
- ❑ None

- ❑ Internal policy, please describe it
- ❑ OpenGIS GML
- ❑ OpengGIS Web Map Context
- ❑ ESRI Shapefile
- ❑ EO format (CEOS, DIMAP…)
- ❑ Other standard, please specify below :_____

➢ Do you follow any standard for the metadata?
- ❑ None
- ❑ Internal policy, please describe it
- ❑ ISO 19115:2003
- ❑ ISO 19115–2
- ❑ ISO 19139
- ❑ INSPIRE compliant
- ❑ Other standard, please specify below :_____

***Curation***

➢ Which policy do you apply for long-term access (migration of data, emulation)? :__
_____
*If documentation is available, please provide a reference :* _____

➢ Which policy do you apply for data reprocessing? :_____
_____
*If documentation is available, please provide a reference :* _____

➢ Which policy do you apply for securing long-term interpretability of data? : _____
_____
*If documentation is available, please provide a reference :* _____

➢ Do you verify and update volatile information (links, references…) within your metadata?
- ❑ Yes
- ❑ No

➢ Do you keep track of volatile information for historical reasons?
- ❑ Yes
- ❑ No

***Upgrade***

➢ Which policy do you apply for archive hardware upgrade and/or substitution?
:_____

➢ Which policy do you apply for archive software upgrade and/or substitution?
:_____

***Data Loss/Discard***

➢ How do you protect your archive against catastrophic events?
:_____

➢ Do you store all the collected data or do you intentionally destroy a part of them? : _____

In the second case, do you have any data selection policy?:

_____

- Did you have some experience in unintentional data destruction? :_____

***Effort***

- What is the effort (man-hours/year) to manage your archive?
  :_____
- How much of that effort (%) is spent in data curation ?
  :_____
- What is the effort (K€/year) to manage your archive?
  :_____

***Sustainability***

- Do you think you can maintain in the future the quality level of your archive (technology, accessibility, standardisation, services)?
  :_____
- Do you think you can increase in the future the quality level of your archive (technology, accessibility, standardisation, services)?
  :_____
- What are the main obstacles to reach these goals?
  :_____
- What are your planned strategies and/or envisaged solutions ?
  :_____

*If documentation is available, please provide a reference :*

_____

# D.2 List of Contacted Organisations

*In the feedback column we will distinguish between partially and fully filled questionnaires, because not all the contacted organisations have provided complete information or they do not have repositories. For the complete answers, a link to the answer is provided in the Organization column.*

✓ : Partially filled
☑ : Fully filled

| COUNTRY | ORGANISATION | FEEDBACK |
|---|---|---|
| Albania | Institute of Forestry and Pasture Research | |
| Andorra | Ministeri de Turisme i Medi Ambient | |
| Austria | ASI-Tirol - Alpines Sicherheits- und Informationszentrum | |
| Austria | Austrian Space Agency | |
| Austria | Austrian Umbrella Organization for Geographic Information | |
| Austria | Bundesamt fur Eich- und Vermessungswesen | |
| Austria | Bundesforschungs- und Ausbildungszentrum für Wald Naturgefahren und Landschaft | |
| Austria | Bundesministerium für Land- und Forstwirtschaft | |
| Austria | Department of Conservation Biology Vegetations and Landscape Ecology | |
| Austria | Environmental Earth Observation Information Technology GmbH | |
| Austria | Geologische Bundesanstalt (Geological Survey of Austria) | |
| Austria | GeoVille | |
| Austria | Graz University of Technology | |
| Austria | Institute of the University of Natural Resources and Applied Life Sciences in Vienna | |
| Austria | International Institute for Applied System Analysis | |
| Austria | INTEGRAL Data Repository Joanneum Research | ☑ |
| Austria | Land Oberösterreich | |
| Austria | Landesvermessungsamt Vorarlberg | |
| Austria | Oesterreichisches Institut für Raumplanung | |
| Austria | Österreichische Raumordnungskonferenz | |
| Austria | Österreichisches Fernerkundungsdatenzentum (Austrian Remote Sensing Data Center) | |
| Austria | Statistik Austria | ✓ |
| Austria | CoreH2O Data Repository – University of Innsbruck | ☑ |
| Austria | Veterinärmedizinische Universität Wien Institut für Medizinische Physik und Biostatistik | |
| Austria | Vienna University of Technology | |
| Belarus | Forest Inventory Republican Unitary Company "Belgosles" | |
| Belgium | Administration générale de la Documentation patrimoniale | |
| Belgium | Agency for Maritime Services and the Coast | |
| Belgium | CRONOS | |
| Belgium | EUROCITIES | |
| Belgium | EuroGeoSurveys | |
| Belgium | Europe AID | |
| Belgium | European Commission Joint Research Center | ✓ |
| Belgium | Eurosense | |
| Belgium | Faculté Universitaire des Sciences Agronomiques de Gembloux | |
| Belgium | Federal Department of the Environment | |
| Belgium | Federal Platform Geographic Information | |
| Belgium | Flemish Institute for Technological Research | |
| Belgium | Geographic Information Management | ✓ |
| Belgium | INFRASIG | |
| Belgium | Institut d'Aeronomie Spatiale de Belgique | |
| Belgium | Management Unit of the North Sea Mathematical Models | |

| COUNTRY | ORGANISATION | FEEDBACK |
|---|---|---|
| Belgium | Ministerie van de Vlaamse Gemeenschap | |
| Belgium | Ondersteunend Centrum GISVlaanderen | |
| Belgium | Région Wallonne - Direction Générale de l'Agriculture | |
| Belgium | Research Institute for Nature and Forest | |
| Belgium | Spacebel | |
| Belgium | Statistik Belgium | |
| Belgium | Universite catholique de Louvain | |
| Belgium | Université de Liège | |
| Belgium | Université Libre de Bruxelles | |
| Belgium | University of Antwerp | |
| Belgium | University of Leuven | |
| Belgium | Urbis | |
| Belgium | Vlaamse Instelling voor Technologisch Onderzoek | |
| Bulgaria | ASDE | |
| Bulgaria | BSDI | |
| Bulgaria | Cadastre Agency | |
| Bulgaria | Ministry of Agriculture and Forestry | |
| Bulgaria | Ministry of Environment and Water | |
| Bulgaria | State Fund Agriculture | |
| Bulgaria | URSIT Ltd | |
| Croatia | Ministry of Agriculture  Forestry and Water Management | |
| Croatia | Sumarski Institut | |
| Cyprus | Cyprus Agricultural Payments Organisation | |
| Cyprus | Cyprus Forestry Department | |
| Cyprus | DLS Data Repository – Department of Lands and Surveys | ☑ |
| Cyprus | NILIS | |
| Cyprus | Oceanography Centre  University of Cyprus | |
| Cyprus | Statistical Service of Cyprus | |
| Czech | CAGI | |
| Czech | Ceske Centrum pro Strategicka Studia  Water Management Department | |
| Czech | Charles University Hydrobiological Station | |
| Czech | Czech Academy of Sciences | |
| Czech | Czech Environmental Information Agency | |
| Czech | Czech Geological Survey | |
| Czech | Czech Hydrometeorological Institute | |
| Czech | Czech Ministry of Environment | |
| Czech | Czech Office for Surveying  Mapping and Cadastre | |
| Czech | Czech Statistical Office | |
| Czech | Forestry and Game Management Research Institute | |
| Czech | Masaryk University  Brno | |
| Czech | Ministry of Informatics of the Czech Republic | |
| Czech | Povodi Lab | |
| Czech | SZIF | |
| Czech | Univerzita Karlova v Praze (Charles University) | |
| Denmark | DAISI (KMS) | |
| Denmark | Danish Centre for Forest  Landscape and Planning | |
| Denmark | Danish Meteorological Institute | |
| Denmark | Danish Ministry of Environment and Energy | |
| Denmark | Danmarks Teknische Universitet | |
| Denmark | Denmark Statistik | |
| Denmark | Department of Agricultural Systems Danish Inst. of Agricultural Sciences | |
| Denmark | Directorate for Food  fisheries and Agrobussines | |
| Denmark | European Environment Agency | |

| COUNTRY | ORGANISATION | FEEDBACK |
|---|---|---|
| Denmark | Geological Survey of Denmark and Greenland | |
| Denmark | NERI-DMU National Environmental Research Institute | |
| Denmark | Risoe National Laboratory | |
| Denmark | Royal Danish Administration of Navigation and Hydrography | |
| Denmark | University of Copenhagen  Inst. of Geography | |
| Estonia | ARIB | |
| Estonia | Eesti Metsamajanduse ja Looduskaitse | |
| Estonia | ELB | |
| Estonia | Environmental Management and Technology Department | |
| Estonia | Estonian Atlas | |
| Estonia | Geological Survey of Estonia | |
| Estonia | Statistical office of Estonia | |
| Estonia | University Tartu | |
| Finland | Centre for Urban and Regional Studies | |
| Finland | European Forest Institute | |
| Finland | FCGI | |
| Finland | Finnish Environmental Institute | |
| Finland | Finnish Forest Research Institute | |
| Finland | Finnish Geodetic Institute | |
| Finland | Finnish Institute of Marine Research | |
| Finland | Finnish Maritime Administration | |
| Finland | Finnish Meteorological Institute | |
| Finland | Geological Survey of Finland  Land use and environment department | |
| Finland | NLS Data Repository - National Land Survey of Finland | ☑ |
| Finland | Soil and Water Ltd | |
| Finland | Statistics Finland | |
| Finland | University of Helsinki Data Repository | ☑ |
| France | ACTeo | |
| France | Agence Unique de Paiement | |
| France | Airmaraix | |
| France | Armines | |
| France | CEA Commissariat à l'Energie Atomique | |
| France | CEMAGREF | |
| France | Centre d'Observation Economique | |
| France | CNES Data Repositories - Centre National d'Etudes Spatiales | ☑ |
| France | Centre National de la Recherche Scientifique | ✓ |
| France | Centre Nationale du Machinisme Agricole du Génie Rural des Eaux et des Forêts | |
| France | CNIG | |
| France | CLS Data Repository- Collect Localisation Satellites | ☑ |
| France | Direction Regionale de l'Environnement Alsace | |
| France | EADS Astrium SAS | |
| France | Ecole des Mines d'Alès | |
| France | ENTENTE & PÔNT | |
| France | EuroGeographics Data Centre | ☑ |
| France | European Space Agency | ✓ |
| France | European-Mediterranean Seismological Centre | |
| France | France Telecom | |
| France | French Institute of Forestry Agricultural and Environmental Engineering (ENGREF) | |
| France | French Research Institute for Exploration of the Sea | |
| France | Générale d'Infographie | |
| France | GeoScience Consultant | |
| France | Infoterra SAS | |
| France | Institut Geographique National | |

| COUNTRY | ORGANISATION | FEEDBACK |
|---|---|---|
| France | Institut National de la Recherche Agronomique | |
| France | Institut National d'etudes Demographiques | |
| France | Institute for Geological and Mining Research | |
| France | Institute Francais de l'Environnement | |
| France | Integrated Global Observing Strategy-Partnership | |
| France | Inventaire Forestier National | |
| France | ISO informatique | |
| France | Laboratoire d`Océanographie de Villefranche | |
| France | Mercator Océan | |
| France | Meteo France | |
| France | Ministère de l'Ecologie et du Développement Durable | |
| France | Ministère de l'interieur Direction de la Defense et de la Securite Civiles | |
| France | Noveltis | |
| France | Office International de l'Eau | |
| France | Paris-Sud University  Systematic Ecology and Evolution Unit | |
| France | Réseau Euro-Méditerranéen d'Information et de FORmation à la gestion des risques | |
| France | SHOM | |
| France | SPOT Image | |
| France | Thales Alenia Space | ✓ |
| France | UNESCO | |
| France | University of Strasbourg Data Repository | ☑ |
| France | Université Pierre et Marie Curie  Paris | |
| France | University of Bordeaux | |
| France | University of Pau | |
| France | University of Rennes | |
| Germany | Albert-Ludwigs-Universität Freiburg | |
| Germany | WDC-MARE and PANGAEA - Alfred-Wegener-Institut für Polar-und Meeresforschung | ☑ |
| Germany | BLOM Deutschland GmbH | |
| Germany | Brandenburg University of Technology Cottbus | ✓ |
| Germany | Bremen University | |
| Germany | Bundesamt für Bevölkerungsschutz und Katastrophenhilfe | |
| Germany | Bundesamt fur Kartographie und Geodasie (BfKG) | |
| Germany | Bundesamt für Naturschutz | |
| Germany | BSH Data Repository- Bundesamt für Seeschifffahrt und Hydrographie | ☑ |
| Germany | Bundesanstalt für Gewasserkunde | |
| Germany | Center for Geoinformation GmbH | |
| Germany | Coordination Center PortalU at the Lower Saxony Ministry of Environment and Climate Protection | ✓ |
| Germany | Delphi Informations Muster Management GmbH | |
| Germany | Deutsche Steinkohle AG | |
| Germany | Deutscher Wetterdienst | |
| Germany | Deutsches Klimarechenzentrum GmbH | |
| Germany | DLR Data Repositories (Deutsches Zentrum für Luft- und Raumfahrt) | ☑ |
| Germany | EADS Astrium Gmbh | |
| Germany | Ecologic Institut fur Internationale und Eurropaische Umweltpolitik gGmbH | |
| Germany | EFTAS Fernerkundung Technologietransfer GmbH | |
| Germany | Entera | |
| Germany | EUMETSAT Data Repository | ☑ |
| Germany | Euromap Data Repository - Euromap Satellitendaten-Vertriebsgesellschaft mbH | ☑ |
| Germany | European Federation of Geologists | |
| Germany | Federal Environment Ministry | |
| Germany | Federal Institute for Geosciences and Natural Resources | |
| Germany | Federal Ministry of Agriculture | |

| COUNTRY | ORGANISATION | FEEDBACK |
|---|---|---|
| Germany | Federal Research Centre for Forestry and Forest Products | |
| Germany | Forschungszentrum Jülich | |
| Germany | Forschungszentrum Karlsruhe GmbH in der Helmholtz-Gemeinschaft | |
| Germany | Fraunhofer Institute for Atmospheric Environmental Research-(IFU) | |
| Germany | Fraunhofer-Institut für Angewandte | |
| Germany | SIB-ESS-C Data Repository - Friedrich-Schiller University Jena | ☑ |
| Germany | Geoforschungszentrum Potsdam | |
| Germany | H. G. Geo Data Solutions GmbH | ✓ |
| Germany | Hamburg Institut fuer Meereskunde | |
| Germany | HYDROMOD Scientific Consulting | |
| Germany | IFM-GEOMAR Leibniz-Institute of Marine Sciences | |
| Germany | Infoterra Gmbh | ✓ |
| Germany | Institut f. Ostseeforschung Warnemünde | |
| Germany | Institute for Tropospheric Research | |
| Germany | Institute for World Forestry | |
| Germany | Koordinierungsstell e UDK/GEIN | |
| Germany | Landesamt für Natur und Umwelt des Landes Schleswig-Holstein | |
| Germany | Leibniz Centre for Agricultural Landscape Research | ✓ |
| Germany | Martin-Luther-University Halle-Wittenberg Inst. of Soil Science and Plant Nutrition | |
| Germany | Max-Planck-Institute for Meteorology | ✓ |
| Germany | Potsdam Inst. for Climate Impact Research | |
| Germany | RapidEye | ✓ |
| Germany | Remote Sensing Solutions GmbH | |
| Germany | Rheinische Friedrich-Wilhelms-Universität Bonn | |
| Germany | Rheinisches Institut für Umweltforschung Universität Köln | |
| Germany | Technical University Hamburg-Harburg Ocean Engineering I | |
| Germany | Technische Universität Dresden IHM-Meteorology | |
| Germany | Technische Universität Dresden Munich Dept. of Soil Science | |
| Germany | Telematica e.K. | |
| Germany | Thuringia State Forest Research Station | ✓ |
| Germany | t-info GmbH | |
| Germany | TuTech | |
| Germany | Umweltbundesamt (Federal Environment Agency) | |
| Germany | Umweltforschungszentrum GmbH | |
| Germany | Univbersitat Triet | ✓ |
| Germany | Universität Stuttgart Institute of Energy Economics and the Rational Use of Energy | |
| Germany | University Bayreuth (UBT) | |
| Germany | University Dortmund Institute of Spatial Planning | |
| Germany | University Hamburg | |
| Germany | University of Duisburg-Essen | |
| Germany | University of Heidelberg Inst. for Environmental Physics | |
| Germany | VCS Aktiengesellschaft | |
| Germany | WDC Marine Environmental Sciences MARUM/AWI Bremen | ☑ |
| Germany | Westfälische Wilhelms-Universität Münster (Institute for Geoinformatics) | |
| Greece | Agricultural University of Athens | |
| Greece | Algosystems SA Research & Development Department | |
| Greece | Aristotle University Thessaloniki | |
| Greece | Association of Municipalities for the Protection of Pendelikos | |
| Greece | Catalan Water Agency | |
| Greece | Greek Biotope/Wetland Centre | |
| Greece | Hellenic Mapping and Cadastral Organisation | |
| Greece | Institute of Mediterranean Forest Ecosystems | |
| Greece | Mediterranean Agronomic Institute of Chania | |

| COUNTRY | ORGANISATION | FEEDBACK |
|---|---|---|
| Greece | NAGII | |
| Greece | National Agricultural Research Foundation | |
| Greece | National Center For Marine Research | |
| Greece | National Kapodistrian University of Athens | |
| Greece | NOA Data Repository (National Observatory of Athens) | ☑ |
| Greece | National Statistical Service of Greece | |
| Greece | Organisation for the Master Plan and Environmental Protection of Thessaloniki | |
| Greece | Siemens | |
| Greece | Space Hellas | ✓ |
| Greece | Space Imaging Europe | |
| Greece | Topographical Service | |
| Greece | University of AEGEAN | |
| Greece | University of Crete | |
| Hungary | Agricultural and Rural Development Agency (ARDA) | |
| Hungary | Central Research Institute for Physics – Atomic Energy Research Institute (KFKI) | |
| Hungary | Eötvös Loránd University Dept.of Meteorology | |
| Hungary | Forest Management Planning Service | |
| Hungary | HUNAGI | |
| Hungary | Institute of Ecology and Botany Hungarian Academy of Sciences | |
| Hungary | Institute of Environmental Management University Gödöllö | |
| Hungary | Institute of Environmental Sciences University of Western Hungary | |
| Hungary | National Institute of Environmental Health-Dept. of Air Hygiene | |
| Hungary | Szent István University of Gödöllö | |
| Hungary | University of Veszprèm - UV | |
| Iceland | LISA | |
| Iceland | Marine Research Institute Reykjavik | |
| Iceland | National Land Survey of Iceland | |
| Iceland | Statistics Iceland | |
| Iceland | University of Iceland | |
| Ireland | Central Statistics Office | |
| Ireland | Coillte Research and Development | |
| Ireland | CSA group | |
| Ireland | Department of Agriculture and Food | ✓ |
| Ireland | Department of Environment Heritage and Local Government | |
| Ireland | Environmental Protection Agency | |
| Ireland | IRLOGI | |
| Ireland | Marine Institute | |
| Ireland | National University of Ireland Galway | |
| Ireland | Ordnance Survey of Ireland | |
| Ireland | Teagasc- Soils and Environment Research Centre | |
| Ireland | Trinity College Dublin | |
| Ireland | University College Dublin Botany Department | |
| Ireland | University College of Cork | |
| Italy | AGEA | |
| Italy | Agenzia per la Protezione dell'ambiente e per i Servizi Tecnici ARPS Piemonte | |
| Italy | ASI Data Repository | ☑ |
| Italy | Associazione per lo Sviluppo della Qualita'Ambientale | |
| Italy | Autonomous Province of Bolzano/Bozen South Tyrol - Forest Department | |
| Italy | Carlo Gavazzi Space SpA | |
| Italy | Centre of Alpine Ecology | |
| Italy | Centro Interregionale CPSG-CISIS | |
| Italy | CESI Ricerca S.p.A. | |

| COUNTRY | ORGANISATION | FEEDBACK |
|---|---|---|
| Italy | Consiglio Nazionale delle Ricerche | ✓ |
| Italy | Consorzio Nazionale Interuniversitario per le Scienze del Mare | |
| Italy | Corpo Forestale dello Stato Servizio CONECOFOR | |
| Italy | ElsagDatamat | |
| Italy | ENEA Data Repositories - ENEA | ☑ |
| Italy | ESRIN - ESA Data Repositories | ☑ |
| Italy | EuroConcepts | |
| Italy | Euroimage | |
| Italy | EURAC Data Repository - European Academy Bozen/Bolzano | ☑ |
| Italy | Federazione delle Associazioni Scientifiche per le Informazioni Territoriali ed Ambientali | |
| Italy | Istituto Geografico Militare | |
| Italy | INGV Data Repository - Istituto nazionale di Geofisica e Vulcanologia | ☑ |
| Italy | JRC Data Repositories - Ispra | ☑ |
| Italy | OGS Data Repositories - Istituto nazionale di Oceanografia e di Geofisica Sperimentale | ☑ |
| Italy | Istituto Nazionale di Statistica | |
| Italy | Italian Space Agency | |
| Italy | KELL | |
| Italy | Ministero dell'Ambiente | |
| Italy | National Agency for Technology Energy and the Environment | ✓ |
| Italy | Planetek Data Processing Centre | ☑ |
| Italy | Politecnico di Milano | |
| Italy | Politecnico di Torino | |
| Italy | Protezione civile | |
| Italy | Remote Sensing Data Engineering | |
| Italy | San Marco Project Research Centre | |
| Italy | Second University of Napoli  Dept. of Environmental Science | |
| Italy | Società Italiana di Fotogrammetria e Topografia | |
| Italy | Tecnomare Spa | |
| Italy | Tele+ Italia S.A.S. | |
| Italy | Tele-Rilevamento Europa | |
| Italy | PREVIEW Data Repositories - Telespazio | ☑ |
| Italy | United Nations Food and Agriculture Organization (FAO) | |
| Italy | Università della Tuscia  Dept. of Forest Environment and Resources (DISAFRI) | |
| Italy | Università di Firenze | |
| Italy | Università di Tor Vergata  Roma | |
| Italy | Universita di Trieste | |
| Italy | University Bologna | |
| Italy | University of Roma Tre | |
| Latvia | Latvian Environment Agency | |
| Latvia | Rural Support Service | |
| Latvia | State Forest Service of Latvia | |
| Latvia | State Land Service | |
| Latvia | University of Latvia | |
| Liechtenstein | Amt für Wald  Natur und Landschaft | |
| Liechtenstein | GDI | |
| Lithuania | Environmental Centre for Administration and Technology | |
| Lithuania | Institute of Geodesy at the Vilnius Gediminas Technical University | |
| Lithuania | Institute of Geology & Geography | |
| Lithuania | Ministry of Environment | |
| Lithuania | National Land Service | |
| Lithuania | National Paying Agency  Control department | |
| Lithuania | State Enterprise “GISCentras” | |
| Lithuania | State Enterprise Centre of Registers | |

| COUNTRY | ORGANISATION | FEEDBACK |
|---|---|---|
| Lithuania | State Forest Survey Service | |
| Lithuania | University of Vilnius | |
| Luxembourg | ACT | |
| Luxembourg | Administration des Eaux et Forêts Service de l'Aménagement des Bois et de l'Economie Forestière | |
| Luxembourg | ASTA-Unité de Contrôle | |
| Luxembourg | SES Astra S.A | |
| Macedonia | University « St. Kiril and Metodij » Skopje Faculty of Forestry | |
| Malta | MRI Data Repository - DG Rural Affairs and Paying Agency | ☑ |
| Malta | National Statistics office | |
| Malta | Plan Making and Policy Development Unit  Malta Environment and Planning Authority | |
| Moldova | State Forest Service | |
| Monaco | Marine Environment Laboratories  Monaco | |
| Netherlands | ALTERRA (Wageningen University and Research) | |
| Netherlands | Energy Research Center of the Netherlands  Dept. Air Quality | |
| Netherlands | European Centre for Nature Conservation | |
| Netherlands | European Union For Coastal Conservation | |
| Netherlands | Geonovum | |
| Netherlands | Ingenieursbureau voor Environmental Analysis and Remote Sensing bv | |
| Netherlands | Institute for Inland Water Management and Wastewater Treatment | |
| Netherlands | Koninklijk Nederlands Meteorologisch Instituut | |
| Netherlands | Marine Information Serve MARIS B.V | |
| Netherlands | Ministry of Agriculture  Nature and Foodquality | |
| Netherlands | National Clearinghouse Geo-Information | |
| Netherlands | National Institute for Coastal and Marine Management/RIKZ | |
| Netherlands | Netherlands Environmental Assessment Agency | |
| Netherlands | Netherlands Geomatics & Earth Observation B.V. | |
| Netherlands | Netherlands Institute for Fisheries Research | |
| Netherlands | Netherlands Institute for Sea Research | |
| Netherlands | Netherlands Institute of Ecology | ✓ |
| Netherlands | Netherlands Space Research Organisation Earth Oriented Science Division | |
| Netherlands | NWO Earth and Life Sciences Council | |
| Netherlands | RAVI | |
| Netherlands | SenterNovem | |
| Netherlands | TNO Data Repository | ☑ |
| Netherlands | University of Groningen  Center for Isotope Research (CIO) | |
| Netherlands | University of Utrecht | |
| Netherlands | Vexcel - Virtual Earth Business Unit | |
| Netherlands | Vrije Universiteit Amsterdam | |
| Netherlands | Wageningen University | |
| Netherlands | Wageningen University  Nature Conservation and plant Ecology | |
| Netherlands | Wetlands International | |
| Norway | Bjerknes Centre for Climate Research  University of Bergen | ✓ |
| Norway | Directorate for Nature Management | |
| Norway | GeoForum | |
| Norway | Institute of Marine Research  Bergen | |
| Norway | KSAT Data Repositories - Kongsberg Satellite Services | ☑ |
| Norway | Miljoverndepartementet | |
| Norway | NERSC Data Repository - Nansen Environmental and Remote Sensing Center | ☑ |
| Norway | Norgit AS | |
| Norway | NILU Data Repositories - Norsk Institute for luftforskning | ☑ |
| Norway | Norsk institutt for skog og landskap | |
| Norway | NINA Data Repository - Norwegian Institute for Nature Research | ☑ |

| COUNTRY | ORGANISATION | FEEDBACK |
|---|---|---|
| Norway | Norwegian Institute for Water Research | |
| Norway | Norwegian Institute of Land Inventory | |
| Norway | Norwegian Polar Institute | |
| Norway | Norwegian research institute | |
| Norway | Norwegian Space Centre | |
| Norway | Norwegian Water Resources and Energy Directorate | |
| Norway | Statens Foruensningstilsyn (Norwegian Pollution Control Authority) | |
| Norway | Statens Kartverk | |
| Norway | Statens vegvesen Vegdirektoratet | |
| Norway | Statistics Norway | |
| Norway | Statkraft | |
| Norway | Storm Weather Center AS | |
| Norway | The Norwegian Meteorological Institute | |
| Norway | University of Tromsø | |
| Norway | UNIVISJON | |
| Poland | AGH University of Science and Technology  Faculty of Physics and Nuclear Techniques | |
| Poland | Agricultural University of Cracow | |
| Poland | Akademia Rolnicza im. Augusta Cieszkowskiego w Poznaniu | |
| Poland | ARMA | |
| Poland | Centralny Osrodek Documentacji Geodeznej I Kartograficznej | |
| Poland | Forest Research Institute | |
| Poland | Geoportal | |
| Poland | Institute of Meteorology and Water Management  Maritime Branch | |
| Poland | Institute of Oceanology  Polish Academy of Sciences | |
| Poland | Institute of Soil Science and Plant Cultivation Pulawy | |
| Poland | Instityt Geodezji i Kartografii | |
| Poland | International Centre for Ecology  Polish Academy of Sciences | |
| Poland | Maritime Institute Gdansk | |
| Poland | NILU Polska Ltd | |
| Poland | Polish Geological Institute Gdansk | |
| Poland | Polish Institute of Environmental Protection | |
| Poland | Poznan University of Technology | |
| Poland | Space Research Centre Polish Academy of Science | |
| Poland | Warsaw University of Technology | |
| Portugal | Direcçao Geral das Florestas | |
| Portugal | INGA | |
| Portugal | INOV | |
| Portugal | Instituto de Investigação Científica Tropical | |
| Portugal | Instituto de Meteorologia | |
| Portugal | Instituto Geologico e Mineiro (Geological Survey of Portugal) | |
| Portugal | Instituto Nacional de Estatistica | |
| Portugal | SNIG | |
| Portugal | Sociedade Portuguesa de Ecologia Faculdade de Ciências da Universidade de Lisboa | |
| Portugal | Superior Technical Institute | |
| Portugal | Universidade Técnica de Lisboa  Superior Inst. of Agronomy | |
| Portugal | University of Aveiro  Departamento de Ambiente e Ordenamento | |
| Romania | Agentia Nationala de Cadastru si Publicate Imobiliara | |
| Romania | Forest Research and Management Institute | |
| Romania | National Institute for Research and Development for Environment Protection | |
| Romania | Romanian Space Agency | |
| Romania | University of Bucharest | |
| RoW-Russia | Centre for Forest Ecology and Productivity  RAS | |
| RoW-Russia | Russian Academy of Sciences | |

| COUNTRY | ORGANISATION | FEEDBACK |
|---|---|---|
| RoW-Russia | Russian Federal Service for Hydrometeorology and Environmental Monitoring (Roshydromet) | |
| RoW-Russia | Russian Remote Sensing Centre | |
| Serbia | Institute for Forestry | |
| Slovakia | APA | |
| Slovakia | Geodesy Cartography and Cadastre Authority | |
| Slovakia | Geological Survey of Slovak Republic | |
| Slovakia | ILE Institute of Landscape Ecology Slovak Academy of Sciences | |
| Slovakia | National Forest Centre | |
| Slovakia | Research Institute of Geodesy and Cartography in Bratislava | |
| Slovakia | Slovak Environmental Agency | |
| Slovakia | Statistical Office | ✓ |
| Slovenia | AAMRD | |
| Slovenia | Gozdarski institut Slovenija Slovenian Forestry Institute | |
| Slovenia | Jozef Stefan Institute | |
| Slovenia | Karst Research Institute ZRC SAZU | |
| Slovenia | SGII | |
| Slovenia | University of Ljubljana | ✓ |
| Spain | Altamira Information | |
| Spain | Autonomous University of Barcelona | |
| Spain | Cartographic Institute of Catalonia | |
| Spain | Centre de Recerca Ecologica i Aplicacions Forestals | |
| Spain | CNIG Data Repository - Centro Nacional de Informacion geografica | ☑ |
| Spain | CIMNE | |
| Spain | Consejo Superior de Investigaciones Científicas | ✓ |
| Spain | European Topic Centre on Terrestrial Environment | |
| Spain | EUSC Data Repository - European Union Satellite Centre | ☑ |
| Spain | Fondo Español De Garantía Agraria (FEGA) | |
| Spain | Forest Technology Centre of Catalonia Laboratory of Plant Ecology and Forest Botany with support of Fundació Territori i Paisatge | |
| Spain | Fundación Centro de Estudios Ambientales des Mediterráneo | ✓ |
| Spain | GMV | |
| Spain | GTD | |
| Spain | IDEE | |
| Spain | Indra Espacio s.a. | |
| Spain | INSA Data Repository - Ingenieria y Servicios Aeroespaciales | ☑ |
| Spain | ICCM Data Repository - Instituto canario de Ciencias Marinas | ☑ |
| Spain | Instituto Español de Oceanografia | |
| Spain | Instituto Geografico Nacional | |
| Spain | Instituto Technológico de Canarias | |
| Spain | Junta de Andalucía | |
| Spain | LATUV | |
| Spain | Ministerio de Medio Ambiente | |
| Spain | Ministry of the Environment | |
| Spain | Parc Científic de Barcelona | |
| Spain | Puertos del Estado | |
| Spain | Sección de Información Meteorológica Instituto Nacional de Meteorología. | |
| Spain | SICE S.A. | |
| Spain | Starlab Data Repository | ☑ |
| Spain | Tecnologías y Servicios Agrarios S.A. | |
| Spain | Telefónica Investigación y Desarrollo | |
| Spain | Universidad Politecnica de Madrid | |
| Spain | Universitat Politecnica de Catalunya | ✓ |
| Spain | University “Jaume I” | |

| COUNTRY | ORGANISATION | FEEDBACK |
|---|---|---|
| Spain | University of Barcelona | |
| Spain | University of Granada | |
| Sweden | Country Administration of Dalarna | |
| Sweden | EuroGOOS | |
| Sweden | FOI | |
| Sweden | Forestry Research Institute of Sweden (SkogForsk) | |
| Sweden | International Geosphere Biosphere Programme | |
| Sweden | Lund University  Dept. of Physical Geography and Ecosystems Analysis | |
| Sweden | Naturvardsverket (Environmental Protection Agency) | |
| Sweden | NLS | |
| Sweden | Stockholm University  Department of Applied Environmental Science (ITM) | |
| Sweden | Swedish Board of Agriculture | |
| Sweden | Swedish Development Council for Land Information | |
| Sweden | Swedish Environment Research Institute | |
| Sweden | Swedish Forest Agency | |
| Sweden | Swedish Meteorological and Hydrological Office | |
| Sweden | Swedish Rescue Services Agency | |
| Sweden | Swedish University of Agricultural Sciences | |
| Sweden | University of Göteborg | |
| Switzerland | CH-CDS | |
| Switzerland | COGIS | ☑ |
| Switzerland | e-geo | |
| Switzerland | Federal Research Institute Forest Ecosystems and Ecological Risks Division | |
| Switzerland | Forschungsanstalt Agroscope Reckenholz-Tänikon ART | |
| Switzerland | Gamma Remote Sensing Research and Consulting AG | |
| Switzerland | INTERLIS | |
| Switzerland | Office Federal de l'Environnement des Forets et du Paysage | |
| Switzerland | Paul Scherrer Institute | ✓ |
| Switzerland | Sarmap s.a. | |
| Switzerland | SIK-GIS | |
| Switzerland | SOGI | |
| Switzerland | Swiss Federal Institute of Environmental Science and Technology | |
| Switzerland | ETH Data Repository - Swiss Federal Institute of Technology Zuerich | ☑ |
| Switzerland | Swiss Federal Statistical Office | |
| Switzerland | Swiss Space Office | |
| Switzerland | UN Environment Programme | |
| Switzerland | University of Bern  Physics Institute  Climate and Environmental Physics | |
| Switzerland | UNOSAT | |
| Switzerland | World Health Organisation | |
| Switzerland | World Meteorological Organisation | |
| Switzerland | World Trade Organisation | |
| Turkey | General Directorate of Forestry | |
| Turkey | Harita Genel Komutanligi | |
| Turkey | Institute of Marine Sciences  Middle East Technical University (IMS METU) | |
| Turkey | MARA | |
| United Kingdom | AEA Technology  Didcot | |
| United Kingdom | Association for Geographic Information | |
| United Kingdom | At-Bristol Ltd | |
| United Kingdom | Avanti Communications Ltd | ✓ |
| United Kingdom | British Geological Survey | |
| United Kingdom | British National Space Centre | |
| United Kingdom | Centre for Ecology and Hydrology | |
| United Kingdom | Centre for Polar Observation & Modelling | |

| COUNTRY | ORGANISATION | FEEDBACK |
|---|---|---|
| United Kingdom | Department of Environment Food and Rural Affairs | |
| United Kingdom | Earthwatch (Europe Office) | |
| United Kingdom | Environment Agency | |
| United Kingdom | Environmental Change Research Centre | |
| United Kingdom | European Centre For Medium-Range Weather Forecasting | |
| United Kingdom | Forestry Commission | |
| United Kingdom | Hadley Centre | |
| United Kingdom | HR Wallingford Ltd Dept. Water | |
| United Kingdom | Imperial College London (UK) | |
| United Kingdom | INFOTERRA Data Repository | ☑ |
| United Kingdom | King's College London | |
| United Kingdom | Lancaster University Department of Geography | |
| United Kingdom | Leeds Metropolitan University | |
| United Kingdom | LGC Limited | |
| United Kingdom | Macaulay Land Use Research Institute | |
| United Kingdom | METREX The Network of European Metropolitan Regions and Areas | |
| United Kingdom | National Assembly for Wales | |
| United Kingdom | National Oceanography Centre Southampton | |
| United Kingdom | National Soil Resources Institute | |
| United Kingdom | National Statistics | |
| United Kingdom | NERC Natural Environment Research Council | |
| United Kingdom | NPA Group | |
| United Kingdom | Ocean Numerics Ltd | |
| United Kingdom | Open Geospatial Consortium (Europe) | |
| United Kingdom | Ordnance Survey | |
| United Kingdom | Ordnance Survey of Northern Ireland | |
| United Kingdom | Plymouth Marine Laboratory | |
| United Kingdom | Rural Payments Agency | |
| United Kingdom | SciSys | |
| United Kingdom | Scottish Natural Heritage | |
| United Kingdom | Staffordshire University Institute for Environment and Sustainability Research | |
| United Kingdom | The Met Office | |
| United Kingdom | UN Environment Programme World Conservation Monitoring Centre UNEP-WCMC | |
| United Kingdom | University of Aberdeen | ✓ |
| United Kingdom | University of Bath | |
| United Kingdom | University of East Anglia | |
| United Kingdom | University of Edinburgh School of GeoSciences | |
| United Kingdom | University of Liverpool | |
| United Kingdom | University of Nottingham | |
| United Kingdom | University of Portsmouth | |
| United Kingdom | University of Reading | |
| Ukraine | Odessa National University Regional Centre for Integrated Environmental Monitoring | |

**Table 3 - Contacted organisations**